codex alimentarius commission

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

WORLD HEALTH ORGANIZATION

JOINT OFFICE: Viale delle Terme di Caracalla 00153 ROME Tel: 39 06 57051 www.codexalimentarius.net Email: codex@fao.org Facsimile: 39 06 5705 4593

Agenda Item 9 **CX/PR 10/42/11**
March 2010

JOINT FAO/WHO FOOD STANDARDS PROGRAMME

CODEX COMMITTEE ON PESTICIDE RESIDUES

42nd Session
Xian, China, 19 - 24 April 2010

DISCUSSION PAPER ON THE CALCULATION METHOD FOR THE ESTIMATION OF MAXIMUM RESIDUE LIMITS FOR PESTICIDES BEING DEVELOPED THROUGH THE OECD

Prepared by the Electronic Working Group Led by the United States of America

Background

At the last 41st session of the Committee, during the discussion of transparency in the maximum residue level estimation process of JMPR, it was noted that "JMPR welcomed a harmonized statistical calculation method and agreed to use the 'OECD calculator for MRLs' as a trial in the 2009 evaluation, if this method was available for the upcoming JMPR meeting"(ALINORM 09/32/24, para. 34). In addition, "To the proposal that JMPR participate actively in the OECD work, the FAO JMPR Secretariat pointed out that JMPR was continually striving for the development and utilization of a statistical calculation method and that JMPR experts actively took part in the development of relevant OECD Guidelines, including the calculation method" (ALINORM 09/32/24, para. 37). Finally, "The Committee also noted that the 'OECD calculator for MRLs' is still in development and there were some questions regarding MRL setting policies which needed to be solved in order to finish this work'(ALINORM 09/32/24,para. 42).

After some discussion, the Committee agreed to send a circular letter containing the questionnaire as presented in CRD 23. The Committee agreed to establish an electronic working group led by the United States and working in English only which would analyze replies and prepare a paper for consideration by the next session of the Committee" (ALINORM 09/32/24, para. 45)

Four responses to the questionnaire were received through the CCPR circular letter. These were from Brazil, Costa Rica, the EU, and Japan. Other responses were received through the OECD including responses from the JMPR. Consideration of this important input along with continuing work by the calculator group has resulted in the calculator group taking a substantially different approach. Release of the details of this new approach is not expected until the beginning of April. However, it is clear that in-depth analysis and discussion of the policy issues relevant to the old version of the calculator, as detailed in the original questionnaire, is not warranted at this time. As a result, this paper provides:

- A summary of all of the responses received to the questionnaire for the information of the CCPR, but with no further analysis
- An update on the current status of the MRL calculator for the information of the CCPR and for their consideration of next steps in relationship to this work

Summary of Responses to the Original Questionnaire

The information provided in Attachment I. is the exact same information that was provided to the OECD. It is provided as two documents (1) a summary of all the responses except from the EU and (2) a separate response from the EU. This was done because the response from the EU was provided after the original summary was completed. The material is provided for your information and no further analysis has been done because the thoughtful and detailed responses, as well as the continuing work by the calculator group, has resulted in a substantially different approach being chosen. These questions are not directly applicable to evaluating the new approach, therefore, an extensive analysis and attempt to reconcile the responses was not warranted.

Update on the Current Status of the MRL Calculator

Currently, it is expected that release of the details of the new approach will be at the beginning of April. An update on the approach now being taken will be provided by a representative of the calculator group at a lunch time or afternoon seminar during the CCPR. As a result of the lack of details currently available on the new approach, it is expected that the discussion at the CCPR meeting will be limited to how members of the CCPR wish to participate in the up-coming review and testing of the calculator as this work proceeds.

The following information is being provided on a very preliminary basis to those who would like to have an idea of what is currently taking place. It is not suitable for comment or discussion at this time, because it is both incomplete and subject to change.

The calculator working group has recently dropped the distributional approaches from the calculator. The calculator working group found, after considerable numerical experimentation that the distributional approaches do behave as expected, increasing their precision as the dataset size increases and providing reasonably good estimates of high percentiles like the 95th for datasets that have at least between 15 and 20 points. But to their surprise they also found out that a version of the "Mean + 3*Standard Deviation" method outperformed the distributional approaches, not just for small dataset sizes as was expected, but also for datasets as large as 20 or 30 points. The specific method now being explored is the "Mean + 4*Standard Deviation" method, where the number of standard deviations taken is increased to four to get a better estimate of the 95th percentile of a lognormal distribution.

The calculator group is also exploring some possible improvements on this method. These can be grouped in the following categories:

- *"Mean + K(n)*Standard Deviation" methods.* As expected with any statistical method, the performance of the "Mean + 4*Standard Deviation" method deteriorates as the dataset size decreases. A possible approach to get a conservative MRL estimate is increasing the multiplier K of the standard deviation accordingly, effectively making it a function of the dataset size n.

- *Minimum Coefficient of Variation (CV) requirement.* For small datasets and for datasets which are highly censored that is, with many values that are at the Limit of Detection (LOD) or Limit of Quantitation (LOQ), the standard deviation may be underestimated. Setting a requirement for a minimum CV (CV = Standard Deviation/mean) may introduce a possible correction to this situation.

- *Censoring*. The "Mean + 4*Standard Deviation" method is robust enough to produce reasonably good results for small and medium levels of censoring. But as the censoring reaches more than 50% of the dataset, its performance deteriorates. How to compute high percentiles of a residue distribution for datasets with high levels of censoring is beyond most existing statistical methodologies. The calculator group is exploring some ad-hoc conservative approaches to this problem, which may lead to the reintroduction of some aspects of the distributional approaches.

Finally, once the proposal for the methodology of the calculator is complete, it will be necessary to return to developing a proposal for rounding. Several approaches remain under consideration at this time.

ATTACHMENT I.

Ad hoc Technical Policy Advisory Group for the MRL Calculator: Summary of Responses to Questionaire

The following summarizes the responses to the questions that the MRL calculator workgroup originally identified and which were sent out to OECD members, JMPR members, and Codex members. Areas of agreement (or near agreement) are identified and where there was disagreement, the issues are highlighted. In some cases follow-up steps are suggested.

Responses were received from: Denis Hamilton & Australian Govt. (Australia), Brazil (description of their MRL setting process with no specific responses to the stated questions), Canada, Costa Rica, CropLife, Karsten Hohgardt & Britta Michalski (Germany), Japan (no specific responses to the stated questions), Netherlands, New Zealand, the United States, and JMPR (compilation of individual's responses with no effort to integrate these; several responses are exactly the same as the ones submitted by individuals/countries-- these were not repeated twice in the summary below).

Question 1. *Does the RCEG/do you have any recommendations as to which of these two competing interests, setting the MRL low enough to discourage misuse vs. setting the MRL high enough to allow GAP to be used without possibility of MRL exceedance, is more important to address when setting MRLs? If so, please carefully characterize and describe the importance of each.*

There seems to be general agreement that the primary consideration should be to set the MRL high enough to prevent exceedances for a pesticide used according to a certain GAP. General support was noted by Australia (Denis Hamilton) (w/caveat that when the MRL is required for control of use, the decision must balance the two considerations), Canada (w/caveat that a balance of the two should be considered to the extent possible), CropLife, Germany (Karsten and Britta), JMPR, Netherlands (w/caveat on the importance of residue trials conducted within 25% of the critical GAP), and the United States (w/caveat that the best method should balance the two).

Costa Rica commented that the setting the MRL low enough to discourage misuse is most important and New Zealand commented judgment should be used on a case-by-case basis to ensure the MRL is set to achieve the appropriate balance between the two requirements.

Australia (Govt.) noted that, in Australia, the use of a statistical calculation method would only be applied to setting MRLs for registration and therefore for the purpose of compliance against GAP and also noted the issues surrounding use of MRLs to estimate dietary exposure for risk assessment purposes which necessitates that best estimates are made.

Of note were the commenters who said that achieving the first interest is not a goal in their MRL setting due to its very nature. This included some members of JMPR which estimates MRLs used as trade standards on the international level and Germany (Karsten) who noted that within the EU the MRL is set based on the worst case GAP. Under extreme circumstances this means that one out of the 27 GAPs of the 27 Member States may drive the MRL while 26 Member States may have a GAP that may tend to give a lower MRL. Using the MRL to enforce misuse is not possible in this case.

Summary/Follow-up: The calculator workgroup would like to provide some illustrations of the implications of choosing very high MRLs, especially for small data sets. This is an example of where we may need to explicitly reframe some of these questions and specifically ask the question in terms of small data sets (versus large datasets), since this is where the problems lie.

Does the RCEG/do you find it acceptable for the MRL to be lower than one or more residues found in a set of supervised field trials when there are many data? If so, how many field trials would be needed to be comfortable with such an MRL? Alternatively, should other non-statistical procedures be implemented to ensure that the recommended MRL is always higher than the Highest Residue?

Responses on this questions were somewhat evenly divided, however, those who did not find it acceptable for the MRL to be lower than the highest residue, seemed to feel very strongly about it. Many of those who found it acceptable for the MRL to be lower than some of the actual residue values noted that this is only an issue where there is a large dataset and did not seem to have as strong an opinion as the other group.

The responders that felt strongly that the MRL should always be higher than any (acceptable) residue found in a field trial included Australia/Denis Hamilton, Canada, JMPR, Netherlands (with the note that the highest residue value is used in the acute consumer risk assessments in the EU and in the dietary burden estimations for livestock feed intake by the EU and JMPR and as long as this is the case it would not be appropriate but noted that if the MRL were used instead of the HR, then there is no objection to setting the MRL lower than the HR), and New Zealand (though not as strongly).

Costa Rica felt it is acceptable; CropLife felt that in most cases the MRL should include the highest residue, but for data sets with more than 19 individual values the MRL could be set lower than the highest residue; Germany (Karsten) felt it would be acceptable but the number of results lower than the MRL should not exceed 5% of the overall number of samples; Germany (Britta) felt that in general the idea of setting MRLs below the highest residue value is not rejected, however, in practice this would only occasionally be done; the United States thinks it is acceptable and would be expected from using a statistical calculation with a large dataset, but notes that its position on this is flexible.

Summary/Follow-up: Given that most actual data sets are small and the strongly held opinions of the group that find it unacceptable for the MRL to be lower than the highest residue, perhaps we can all agree to go with that conclusion. However, this is another example of where we may need to explicitly reframe the question and specifically ask the question in terms of the size of the data set.

It might be useful for those who feel strongly about the MRL always being higher than the highest residue, to answer the following question-- Is there any size data set that would be large enough to make an exception to this? If so what size?

Question 2. *In order to determine the minimum sample size appropriate for calculating MRLs, could the RCEG/you please specify the target percentile(s) and characterize how much variability is acceptable in an MRL estimate?*

Most responders mentioned the 95th and or 99th percentiles in their responses. These included CropLife which supports the 95th percentile; Germany (Karsten) supporting not less than 95th percentile and noting that a political goal proposed in the past was to reduce the MRL exceedances to less than 1% of samples and to reach this goal, it would be helpful to set the target percentile at least at the 99th percentile; Netherlands supporting the lower bound of a given percentile noting that to compensate for the expected lower values of this criterion, a higher percentile could be chosen, e.g., instead of taking the upper bound of the 95th percentile, take the lower bound of the 99th percentile, or it might be considered to use a lower confidence level, if needed; and the U.S. supporting somewhere between the 95th and 99th percentiles and proposes that the minimum sample size be one of the last policy decisions determined and that it be based on simulations incorporating the final (or nearly final) statistical methods proposed for use.

Australia (Govt.) suggested that a statistical method for estimating MRLs would be adequate if the 2.75th percentile MRL outcome is larger than the true 90th percentile residue (rounded) of the underlying population and the 97.5th percentile MRL outcome is less than the rounded 99.9th percentile of the underlying population.

Australia/Denis Hamilton noted that small data sets are a fact of life and there should be no expectation that statistical calculations will be helpful because uncertainty in the results of statistical calculations on small data sets will be so wide as to make their results of little use. Canada supported, based on the work done using real data sets and the resulting proposed MRL values, that a minimum sample size that will result in "reasonable/realistic" MRL values be used.

Question 3. *Does the RCEG/do you have any recommendations for proposing MRLs when there is concern about excessively high MRL proposals for small datasets?*

Does the RCEG/do you consider the use of a regulatory ceiling appropriate? If so, does the RCEG agree with the currently defined ceiling of the maximum of 2 X HR or 3 X STMR?

Many supported the idea of a regulatory ceiling, though not necessarily the one currently proposed.

Canada: In response to the first question, an option for a small data set would be to default to the mean + 3SD approach. It is suggested that, integrated into the MRL calculator guidance document, is a section that deals with small data sets for which the calculator is not appropriate. An approach must be developed and become part of the guidance for calculating MRLs. Concerning the second question, Canada does not recommend a defined regulatory ceiling like maximum of 2 x HR or 3 x SMTR for datasets where the MRL calculator can be used. However it may be worth exploring the use of a regulatory ceiling for small data sets.

Costa Rica: It is appropriate to use a regulatory ceiling but it must be justified scientifically.

CropLife: Support regulatory ceiling approach which they believe describes an important element of the "expert judgment" which is used by regulatory authorities.

Germany: Supports defining an upper limit or ceiling (requiring more data is not successful at the moment). The EU does not currently have a regulatory ceiling. Experience shows that MRLs are rarely higher than 2 X the HR or 3 X the STMR. Nevertheless, concerning public perception the ceiling alone is not helpful. There exists an additional need to discuss (and review) the IESTI/NESTI calculation.

New Zealand: Scientific judgment and experience should be used to moderate/refine/adjust any suggested "extreme MRL values" derived by the calculator from limited data sets. There is merit to use of a regulatory ceiling limit.

Netherlands: Does not agree with a ceiling—it would be more appropriate if the calculator gave a warning that this ceiling of 2 x HR or 3 x STMR is exceeded. It would then be up to the risk managers to decide whether to use the statistical value or another value. It is important that the MRL calculator is based on sound statistical calculations and not on arbitrary criteria.

Australia (Denis Hamilton & Govt): It is too much to expect that the 2 X HR or 3 X STMR would be suitable for all situations and all data sets. Australia (govt) also noted that imposition of arbitrary constraints on possible MRL outcomes is questionable.

U.S.: The US does not consider the proposed regulatory ceiling of the maximum 2 X HR or 3 X STMR to be appropriate. US promotes use of percentiles (and/or confidence intervals) for setting MRLs because such statistical estimates are much less variable than the highest (observed) residue. Although the median residue is much more stable, the US prefers to derive a relationship between the median residue and a (fixed) upper percentile residue when determining the appropriate facto to multiply by the median, when certain assumptions are made with regard to the distribution of filed trial residues.

Summary/Follow-up: Questions 2 and 3 are, perhaps, the most difficult questions we have to answer. Domingo has informed me that the calculator group is working on these questions for small data sets and there will soon be more information for the group to consider. Do people think there is merit to the US suggestion that this be the last policy question to answer—since simulations would be run with the most "final" version of the calculator available? Do people think there is merit to the Canadian suggestion that, integrated into the MRL calculator guidance document, should be a section that deals with small data sets?

Question 4. *Does the RCEG/do you agree with the Calculator Workgroup's proposal that a fully automated procedure be used for selecting the most appropriate statistical distribution?*

Most felt that a fully automated procedure for selecting the most appropriate statistical distribution was acceptable. This included Canada (with the caveat that the selection process is transparent since the evaluators will need to defend the statistically-based decisions in the absence of complex statistical knowledge; CropLife; Germany (Karsten) (with the caveat that a prerequisite for a fully automated

procedure is the evaluation of the procedure, i.e. that the calculation method used gives correct results) and Germany (Britta) (with the caveat as long as the outcome of the rejected approaches is reported); most JMPR responders (with the caveats that the procedure must be shown to work and the user must be able to view all of the outputs necessary to determine that the selected distribution was a good fit); New Zealand; Netherlands (with the caveat that the data are transparent and both the Goodness of Fit and the probability plots are shown); the US (with the caveat that it is inappropriate that no "after-the-fact" judgment can be applied **to the MRL** which would include consideration of whether the selected distribution was appropriate).

Costa Rica felt it is not acceptable, rather the user should select the most appropriate distribution.

New Zealand "has some concerns" about a fully automated procedure.

Australia (Denis Hamilton and Govt.) seemed opposed to the idea noting that the operator needs to understand what is happening and to observe if assumptions, extrapolations, etc. are reasonable. For some people, a fully automated procedure may promote calculations on inappropriate data sets.

Summary/Follow-up: Does anyone think it is *not* acceptable to have a fully automated procedure that runs, produces the results--including all necessary documentation-- which the user can then use to make an independent evaluation and subsequently re-run the program with changes that might include the user selecting an alternative distribution? The user guidance document would require the user to document the reasons for all changes made, for example, the selection of a different distribution.

Question 5. *Does the RCEG/do you find it acceptable to reject outliers based solely on statistical tests? If not, what extenuating circumstances would warrant the removal of observations viewed as excessively high? Would the RCEG support the use of a statistical outlier test in conjunction with extenuating circumstances when determining whether or not to classify an observation as an outlier?*

No responder supported rejecting outliers based solely on statistical tests although most supported using statistical tests to identify observations that needed to be reviewed in detail. Most responders noted that removal of observations was acceptable in cases where the observation appeared to be a statistical outlier *and* there were doubts concerning the validity of the observation after detailed review of the study results.

Several respondents distinguished between small and large data sets, noting that for small data sets it is not possible to identify outliers.

Summary/Follow-up: Can we agree that it is appropriate to run statistical tests to help identify possible outliers and then instruct the user to do additional research on the circumstances surrounding the observation? The guidance document would instruct the user that an observation should only be removed when there is evidence from the study results that the observation may not be valid. The user would be required to document what this evidence was.

Question 6. *Does the RCEG/do you consider the system of MRL classes proposed in the draft calculator to be an appropriate compromise (between the less fine EU method on the one hand and the more fine NAFTA method on the other)?*

Most responders can accept the proposed MRL classes. Canada says they appear to be reasonable; Costa Rica supports, CropLife supports; Germany (Karsten) supports; Germany (Britta) supports in practice (with caveat that given the uncertainties MRL classes differing from each other by less than a factor of 2 (e.g. more classes than 0.1, 0.2, 0.5, 1… cannot really be distinguished); New Zealand could support (with caveat that they would prefer for the value of each class to be about 50% higher than the value of the previous class and strongly favor deletion of the values 0.015 and 0.15 as they do not appear to add value); U.S. supports.

Netherlands does not consider the proposal acceptable and argues that MRLs should be rounded to one figure (0.01, 0.02,…., 0.1, 0.2,…., 1, 2,…, 10, 20,…100, 200,…) due to the variability of the residue data.

Australia (Denis Hamilton & Govt) indicates that a full assessment of the calculator uncertainties be conducted and this would suggest suitable steps in the MRL classes.

Does the RCEG/do you support the Calculator Workgroup's proposal to round down MRL calculations in very limited cases?

Many found the proposal acceptable including Canada, Costa Rica, CropLife, Germany (Karsten), Netherlands, New Zealand, United States.

Australia (Denis Hamilton & Govt.) felt this question should be re-examined when the values slightly above an MRL class have their uncertainties attached.

Germany (Britta) noted that the differences resulting from rounding are much smaller than the variations in the datasets and it is up to management to decide whether to round up or down of if mathematical rounding is used in all cases.

United States noted the classes are not defined or otherwise limited by units therefore lower MRLs can be automatically accommodated by simple expressing the concentrations in a different unit than mg/kg.

One JMPR commenter was opposed to the proposed rounding scheme and favoured rounding using standard rules. In the example values <0.41 are rounded down to 0.4, values <0.45 are rounded down to 0.4, and values > or equal to 0.45 are rounded up to 0.5 but also noted it was a somewhat academic discussion since Crop Field Trial data is only significant to two figures and asked the question if the rounding should be done before entering the data into the calculator or after the calculation?

Would the RCEG/you be in favor of setting lower MRL classes such as 0.001, 0.002, and 0.005 mg/kg as demonstrated by limit of quantitation (LOQ) of the enforcement analytical method for a specific pesticide/commodity combination? Does the RCEG consider an adjustable case-by-case lowest MRL class acceptable or does the RCEG have other specific guidelines for establishing low MRL classes?

Most responders were in favor of including lower MRL classes. Australia (Denis Hamilton & Govt.) noted that Codex MRLs are already established at such levels as 0.003 and 0.005 and have been for many years, therefore, there should be no reason to exclude them; Canada supports setting MRL classes lower than 0.01 (otherwise discourages registrant from developing better methods) and believes that specific guidance should be provided otherwise the process will be very subjective; Costa Rica supports establishing lower MRL classes if justified by the limit of quantitation of the analytical method; Germany (Karsten) notes that currently there are a few cases with MRLs below 0.01 (organo-chlorine compounds in products of animal origin) and in the near future they will discuss setting MRLs below that value—which will be a case-by-case decision also these low values will cause problems to explain that such low residues are not part of a misuse); New Zealand support lower MRL classes if there are finite residues and analytical methods available to measure them.

Netherland felt it is not appropriate because you get in a never ending loop with the analytical laboratories being forced to invest in ever more expensive equipment to no purpose (with the caveat that a lower MRL might be required e.g. for a very toxic compound—so there must be a possibility to choose for lower MRL classes like 0.001—but it should be documented why such a low MRL is required and it might be questioned why such a toxic compound is authorized).

Croplife felt it is not appropriate to use smaller MRL classes because MRLs need to be internationally enforceable by multi-residue methods which are not this sensitive.

One JMPR commentor felt the classes should be set at one significant figure, and the lowest class should be determined by the demonstrated LOQ in each case and another found both questions "immaterial".

Summary/Follow-up: There seems to be quite a bit of agreement in these areas. Perhaps the responders who do not agree with the majority could comment on how strongly they feel about their positions.

11/11/2009

European Community comments on the Codex Circular Letter CL 2009/19-PR: Request for Comments on the MRL Calculation Method being developed through the OECD

1. Setting MRLs "high" or "low"?

Question 1: *Which of these two interests is more important to address when setting MRLs?*

Within EC the MRL (fixed at EC level) is set based on the worst case or critical GAP (GAPs are defined at national level). This means that in some cases one out of 27 GAPs of the 27 EC Member States may drive the MRL while 26 MS have a GAP that would lead to a lower MRL. In the case of Import Tolerances or MRLs based on Codex MRLs, the critical GAP is applied outside the EC and the MRL may reflect residues of pesticides that are not even authorised in the EC.

This shows that the MRL is not suitable to enforce correct application of all the nationally authorised GAPs. To enforce the correct GAP, additional measures are taken including controls at the farm gate or on the farm.

Nevertheless, for reasons of consumer protection, the EC MRLs are set low enough that using more product than prescribed in the critical GAP will likely result in an MRL exceedance.

The value proposed is always going to be a compromise between these two key aspects. Whereas an MRL to accommodate GAP is the ultimate aim, there will always need to be a compromise if interest i)(*the MRL should be set low enough to discourage misuse, such that using more product than prescribed by the GAP will likely result in an MRL exceedance*) is also considered. The current approach taken in the EC is to try to minimise the GAP i.e. set the GAP as low as possible considering all other aspects of the risk assessment and the efficacy of the use rate. If this approach is taken then in theory setting the MRL high enough to accommodate the minimised GAP should also meet the criteria of interest i).

One aspect that is critical to the approach of setting the MRL high enough such that use according to GAP is unlikely to result in an MRL exceedance is to quantify the term “unlikely”.

In the EC it is agreed that such an exceedance should not be more than 5%. Interestingly, the results of the EU monitoring programme show that since 1996 the violation rate fluctuates between 3 and 5 %, but part of this.

Neither of the two aspects prevail in the EC; they are in a tight balance. To be able to strike the right balance the EC requires a large enough dataset.

Question 2: *Is it acceptable for the MRL to be lower than one or more residues found in a set of supervised field trials when there are many data? Alternatively, should other non-statistical procedures be implemented to ensure that the recommended MRL is always higher than the HR?*

From a statistical point of view, setting of an MRL at a level lower than the respective HR is acceptable in case of large datasets. In fact, up to 5% of the overall number of samples in one dataset may exceed the MRL because of the present methodology.

Nevertheless, it must be kept in mind that in many cases datasets which are available for the estimation of MRLs include 8 or less uncensored data points depending on the importance of the crop (in Europe a data set of 8 (major crop) or 4 (minor crop) trials per region within the EU is defined as a MINIMUM). These rather small datasets normally result in an underestimation of the high percentiles used for MRL setting (compared to the underlying parent population). Under practical conditions the usage of MRLs being lower than the

highest residues found in field trials which match the critical GAP should be considered carefully on a case by case basis.

2. Minimum dataset needed for a statistical technique for MRL setting

Question 3: *In order to determine the minimum dataset size appropriate for calculating MRLs, how much variability is acceptable in an MRL estimate?*

The accuracy of MRL estimates should be higher for commodities that are major contributors to dietary intakes (chronic and acute). In addition, when the MRL is closer to the highest toxicologically acceptable level, less variability is acceptable.

To assure that above goal is met the target percentile should not be less than 95th percentile. However, the overall goal should be that MRL exceeding should be well below 5%.

The acceptable variability of the residue data strongly depends on the use being evaluated and is therefore difficult to generalise. For "normal" applications like foliar spray applications on cereals or fruits, a variability of 2-3 orders of magnitude is normally accepted within the residue data. In special cases like granular applications, an even higher range of results is acceptable, since normally a nil-residue situation occurs with a few exceptions of very high residues. No real "distribution" is present in these cases.

The variation in MRL estimations should still allow determining which MRL class is most appropriate to cover residues from the intended use. Normally comparable datasets should result in the same or a closely related (next highest or next lowest) MRL class. In general, high residues found in field trials should have a higher impact on the estimation of the appropriate MRL class than low residues detected near the LOQ.

Question 4: *Are there any recommendations concerning how to set MRLs when there is concern about excessively high MRL proposals for small datasets?*

The value of an MRL derived from field data will in any procedure depend on the number of available data.

A simple procedure like rounding-up the largest residual has the disadvantage that it is very sensitive for single data points (possibly outliers).

A statistical procedure where the upper bound of a given high percentile is taken has the advantage of being less sensitive to single data points, but it has a very serious drawback: the smaller the sample size, the higher the upper bound of that percentile.

This problem is "solved" in the MRL calculator by proposing the 95/99 rule. This is an *ad hoc* solution, which undermines the starting point of trying to quantify the uncertainties (limitations) in the data. Besides, it might be quite advantageous for the manufacturer. The manufacturer may just start with the lowest number of data required, hoping that the upper bound of the 95th percentile will be higher than the point estimate of the 99th percentile. In addition, there is a fifty percent probability that even the 99th percentile is overestimated. The point estimate of the 99th percentile might even be smaller than the (real) 95th percentile in a particular case.

Instead, it could be considered to establish the MRL based on the lower bound of a given percentile. To compensate for the expected lower values of that criterion, a higher percentile could be chosen. E.g., instead of taking the upper bound of the 95th percentile, take the lower bound of the 99th percentile. Or it might be considered to use a lower confidence level, if needed. This procedure will "punish" for a situation where few data are available. At least, it is a straightforward procedure, which avoids difficult discussions on "special" cases, and temptations to delete single observations that make the 95/99 procedure more profitable. In addition, the use of lower confidence bounds makes the issue of the "cut-of" sample size less critical.

3. Introduction of an upper limit for the MRL (Regulatory ceiling)

Question 5: *Is the use of a regulatory ceiling appropriate? If so, is the currently defined ceiling of the maximum of 2 x HR or 3 x STMR appropriate?*

In the EC a regulatory ceiling is not applied. The experience shows that MRLs are hardly ever higher than two times the HR or three times the STMR (the last point can be estimated from the EC method I when taking into account that the mean is in the same order of magnitude than the standard deviation.)

We feel it would be more appropriate if the calculator would give a warning that this ceiling of 2x HR or 3x STMR is exceeded. It would than be up to the risk managers to decide whether to use the statistical value from the MRL calculator or another value. But it is important that the MRL calculator is based on sound statistical calculations and not on arbitrary criteria.

The EC believes that the basis for an MRL proposal should be statistical analysis of the data available. A certain amount of expert judgment can be applied if the value seems unreasonable high based on the data set available, however if there is a significant amount of variability within the results of a data set then this would indicate good reason to set the MRL at a higher value (it is more **likely that variable residues occur and so the MRL needs to be higher to accommodate this).**

4. Transparency of expert judgement in MRL setting

Question 6: *Is it acceptable that a fully automated procedure be used for selecting the most appropriate statistical distribution?*

A fully automated procedure is acceptable if the calculator is used as a first step in decision making. This would be better for transparency because when in later steps case by case decisions are taken based upon expert judgement this needs clear justification.

However, a prerequisite for a fully automated procedure is the transparency on the modules used and a transparent evaluation of the procedure, i.e. to guarantee that the calculation method used gives correct results (Application of "Good Computer Practice").

It will also depend on how the statistical fit is determined. The EC has some concerns over the approach of choosing the "best fit" from a selection of distributions given that the data sets used are quite small. The uncertainties around the different possibilities for the fit need to be clear (perhaps by some sort of weighting for each one –this would of course require expert judgment to interpret the data and reach a logical conclusion) so that the overall approach used is transparent.

The methodology used in the statistical calculators needs to be clearly and transparently explained so those using the calculator can clearly explain the uncertainly around the results to risk managers.

5. How to treat outliers in residue data set

Question 7: *Considering the differing views of OECD member countries regarding the use of outlier tests, is it acceptable to reject outliers based solely on statistical tests?*

Statistical tests for outliers should be used very carefully. Rejection of a value as being an outlier on the basis of a statistical test solely is not appropriate. A concerned residue study should be examined carefully and if there are doubts that the study was conducted correctly the value could be discarded.

In the EC outlier tests are used only at the upper end of the distribution denying that outliers could also occur at the lower end of the distribution.

Rejecting the outliers is only acceptable when justification (biological or experimental) is provided and not only on the basis of statistical tests. The number of results generally available from the residue trials is not large enough for robust statistics, without any justification the EC does not know really the difference between an extreme value or an outlier. One could ask up to which percentile (99%, 95%) we can consider the extreme values for the evaluation.

6. Rounding and MRL classes

Question 8: *Is the system of MRL classes proposed in the draft calculator an appropriate compromise?*

Results from one set of trials according to GAP have a quite high variability; furthermore measurement uncertainty is not yet defined. Too narrow classes give a false impression of precision.

MRLs should be rounded to one figure (0.01, 0.02, …., 0.1, 0.2, …, 1, 2, …, 10, 20, …, 100, 200, ….). Note one figure is not one decimal! Therefore the 0.015, 0.15 and 1.5 values are not appropriate. The variability of the residue data does not allow such tight limits and rounding to one figure is appropriate. If you look only at analytical variability than an RSD of 20% is allowed RSD means 1x standard deviation and generally a value of 2 stdev is considered appropriate to indicate the confidence interval around the mean. This means that the value that you measure, e.g. 1.1 could as well be 0.66 (1.1-2x20%) or 1.5 (1.1+2x 20%) or anywhere in-between. But if you also want to take in homogeneity of sampling into account than the confidence intervals become even larger. So there is no point in setting an MRL at more than one figure.

Although a statistical calculation is used to estimate an MRL from the selected residue values, the variability around these selected values is not taken into account. You do not gain precision from statistical calculations, if this variability is not taken into account. The MRL calculator should be used to get a minimum MRL value.

Too small classes may lead to problems due to analytical uncertainty. Under practical conditions a difference in residue levels of less than 25 % can not be distinguished due to analytical uncertainty. Taking also into account intra-laboratory variation, a difference in residue levels of less than 50 % can not be distinguished. Nevertheless, with 25% level in uncertainty the proposed classes are barely acceptable.

Question 9: *Is the Calculator Workgroup's proposal to round down MRL calculations in these cases acceptable?*

It is not clear why the conventional (international) approach to rounding could not be used in this instance i.e. for the example given calculations of 0.449 and below would be rounded down to 0.4 and calculations of 0.450 and above would be rounded up to 0.5. Going back to the first question if the aim is to set the MRL *high* enough such that use according to GAP is unlikely to result in an MRL exceedance the rounding up should always be employed. Either way it should be made clear what approach is being taken to ensure consistency.

Question 10: *Should lower MRL classes such as 0.001, 0.002, and 0.005 mg/kg be included as demonstrated by limit of quantitation (LOQ) of the enforcement analytical method?*

There will always be need to be a balance between setting the lowest class to reflect current analytical capabilities and allowing for quicker (and cheaper) monitoring of residues in food. The EC is aware of these issues in relation to the implication of the current EU MRL Regulation which has in place a "default" LOQ value of 0.01mg/kg for all MRLs/uses not notified. This is proving challenging for our monitoring laboratories given that historically higher levels were considered suitable LOQs for monitoring purposes. Consideration should only be given to lower MRL values where the toxicity of the compound requires it, or where there are specific risk issues requiring lower levels.

For legal certainty it is good to indicate when an MRL coincides with the generally agreed LOQ for routine laboratory analysis for a particular combination.

Question 11: *Would an adjustable case-by-case lowest MRL class be acceptable or should we consider other specific guidelines for establishing low MRL classes?*

MRL should be set as low as necessary from a toxicological point of view. If necessary even lower than 0.01 mg/kg, but if there is no toxicological necessity the MRLs should be as low as feasible but not below the value of 0.01 mg/kg if the consumer risk is acceptable.