

# 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 - E-mail: [codex@fao.org](mailto:codex@fao.org) - [www.codexalimentarius.org](http://www.codexalimentarius.org)

Agenda Item 17

CRD07

April 2024

ORIGINAL LANGUAGE ONLY

## JOINT FAO/WHO FOOD STANDARDS PROGRAMME

### CODEX COMMITTEE ON CONTAMINANTS IN FOODS

17th Session  
Panama City  
15-19 April 2024

#### REPORT OF THE PRE-SESSION WORKING GROUP ON THE GUIDANCE ON DATA ANALYSIS FOR DEVELOPMENT OF MAXIMUM LEVELS AND FOR IMPROVED DATA COLLECTION

*(Prepared by European Union as Chair of the WG on the Guidance on data analysis for development of maximum levels  
and for improved data collection)*

#### INTRODUCTION - BACKGROUND

1. CCCF16 (2023)<sup>1</sup> agreed
  - (i) on the proposed changes to the GEMS/Food database as presented in Appendix I of CX/CF 24/17/17;
  - (ii) on the workplan for the coming year for the section “Data collection, data submission and data extraction”, i.e., after feedback by the GEMS/Food database administrator on which of the recommendations can be effectively implemented and on the timeframe of their implementation, the section “data collection and submission and data extraction” would need to be updated taking into account the feedback from GEMS/Food database administrator. The updated section would be circulated for comments to the EWG and finalised for submission to the Codex secretariat for circulating for comments in view of finalisation of this section at CCCF17;
  - (iii) on the topics to be addressed in the sections “Data selection/clean-up – generating overview of data” and “statistical analysis”;
  - (iv) on the proposed workplan for the coming year on the sections “Data selection/clean-up – generating overview of data” and “statistical analysis”, i.e. to update the sections “Data extraction/selection/clean-up – generating overview of data” and “statistical data analysis” containing the basic elements and principles and to circulate the updated sections to the EWG for comments and to submit the outcome of the EWG consultation to the Codex secretariat for circulating for comments in view of a possible provisional agreement at CCCF17.
  - (v) that a list of topics of sections “Data selection/clean-up – generating overview of data” and “statistical analysis” shall be elaborated for consideration and agreement by CCCF17 for further discussion after CCCF17;
  - (vi) on the conclusions as regards:
    - (a) the provisional minimum number samples for a 95<sup>th</sup> percentile estimation with 95% confidence;
    - (b) whether a combined dataset or individual datasets should be used for developing MLs,
    - (c) further consider the role of the Committee in calculating dietary exposure reduction rates when considering MLs.

---

<sup>1</sup> REP23/CF16 para 98

- (vii) to recommend to WHO the development of additional training materials and opportunities for the data submission to and data extraction from the GEMS/Food database and to recommend the Codex Member countries to provide the necessary funds for this;
- (viii) on a more structured process for elaborating calls for data;
- (ix) on the consideration of data availability and quality before deciding on new work;
- (x) to holding of a physical meeting of the WG immediately prior to CCCF17 to discuss the guidance document; and
- (xi) to re-convene the EWG chaired by the European Union, co-chaired by Japan, the Netherlands, and the United States of America, working in English, to continue the work on a proposal for a general guidance on data analysis for ML development and improved data collection

#### **WORK PERFORMED/NOT PERFORMED SINCE CCCF16**

2. The GEMS/Food administrator was much too late consulted by the Chair of the EWG and therefore the section "Data collection and submission" could not be updated taking into account the feedback from GEMS/food database administrator and be circulated for comments to the EWG and submitted to CCCF17 for finalisation. The section "data collection and submission and data extraction" of the guidance document, as it currently stands, is provided in Appendix II to CX/CF 24/17/17. This part needs still to be updated once the feedback from the GEMS/Food Administrator on the feasibility and acceptance of the recommended changes by CCCF16 has been received.
3. The sections "Data selection/clean-up" and "data analysis" were not updated by the Chair of the EWG and not submitted for comments to EWG. Therefore, no updated document of the sections "Data selection/clean up – generating overview of data" and "statistical data analysis" is submitted to CCCF17 for discussion and possible provisional agreement.
4. A list of headings/topics to be addressed in the sections "Data extraction/selection/clean-up" and "data analysis" for possible discussion at CCCF 17 is provided in Appendix III of CX/CF 24/17/17.
5. Appendix IV of CX/CF 24/17/17 contains for information only the sections "Data extraction/selection/clean-up" and "data-analysis" as presented in Appendix IV of CX/CF 23/16/12, in which the outcome of the discussions at CCCF16 has been integrated.

#### **ENVISAGED WORK PROCEDURE**

6. As the work on the draft Guidance on data analysis for development of maximum levels (MLs) and for improved data collection has not progressed as initially foreseen, due to the inactivity of the Chair of the EWG, another work procedure is proposed for discussion at CCCF17.
7. As the guidelines on data analysis are being developed for internal CCCF working procedure, it is proposed to convert this work from an EWG into a pre-session working group (which could operate in a physical or virtual mode) or an in-session working group (hereafter referred to as "the WG") similar to the work of priorities for /follow-up to JECFA and the review of standards.
8. The envisaged work procedure will be discussed and agreed at the Plenary session of CCCF17.

#### **POINTS SCHEDULED FOR DISCUSSION AT THE PRE-SESSION WORKING GROUP**

9. Discussion on the preliminary feedback from the GEMS/Food database administrator (full feedback not available given the need to discuss certain issues with the developers)
10. Discussion on the topics for the sections "Data selection/clean-up" and "data analysis" as provided in Appendix III of CX/CF 24/17/17 (in particular the merging/combining of certain topics from the section "data analysis" with topics of the section "data selection/clean-up").
11. Checking of completeness of the topics identified and listed in Appendix III for which further discussion is needed in the future (i.e., post CCCF17)
12. Confirmation of the correctness of the integration of the outcome of the discussions at CCCF16 into the sections "Data selection/clean up" and "data-analysis".
13. (Ev.) the level of detail/complexity of certain aspects (in particular as regards certain parts of the data analysis).

**OUTCOME OF THE DISCUSSIONS AT THE PRE-SESSION WORKING GROUP**

14. Due to time constraints, only the preliminary feedback from the GEMS/Food database administrator as provided in Annex I to this report could be discussed. It concerns a preliminary feedback as full feedback was not available due to the late request to the GEMS/Food database administrator and the need to discuss certain issues with the GEMS/Food database developers.
15. On the preliminary feedback provided by the GEMS/Food database administrator following points were discussed in the pre-session Working Group:

**Existing fields**

- a) Field G "Country/region": OK to add flag and not to change the field name although the combination changing field and adding flag would be preferable.
- b) Field N: "Measurement units for Contaminant Levels": it was stressed that the proposed change is not a cosmetic change as the data submitter must be fully aware that the measurement unit is per kg (and e.g., not per g). As the data extracted are provided in measurement unit per kg it is very important that the measurement units in the drop-down menu are also mentioned as per kg.
- c) Field O "LOD" and P "LOQ": The requested order change does provide added value as the LOQ field would become mandatory for all newly generated /submitted data. The field O "LOD" would then be optional for all newly generated /submitted data. It is stressed that due to then processed changes to the fields the previously submitted data remains valid and are to be used in the discussion on the development for MLs even if certain fields that have become mandatory were not filled for these data.
- d) Field Q: "Results based on" it was highlighted that the proposed change (addition of "as sold" to the drop-down choice "As is") would not have an effect on the legacy data. Agree to introduce information of fat/water content in new proposed field "Compositional information".
- e) Field R "Portion Analyzed": the proposed change of "total" into "whole" provides more clarity and is not considered a cosmetic change.
- f) Field S "State of Food Analysed" the proposed addition of "(cooked, raw)" to the field name is not a cosmetic change but provides clarity to the data submitter as the data submitter sees immediately what kind of information has to be provided in that field without having to open the drop-down menu and the addition shall avoid information provided in the wrong field.
- g) Field V "Confidentiality". The proposed change (yes/no) provides more clarity than the current "confidential yes or blank". In case the proposed change would not be accepted it could be appropriate to add a flag that all data for which "blank" is chosen will be considered as non-confidential in data handling and analysis. It was furthermore confirmed that confidential data would be available to the EWG for the development of the ML, but possibly certain aspects (e.g., country of origin) might not be disclosed. The importance of involving the GEMS/Food database administrator in any data extraction for discussion in the EWG was stressed.

**New fields**

It was highlighted that all new proposed fields are considered to provide added value based on recent experiences on analysing GEMS/Food database data for ML development.

- h) New field: "Year of production": This is relevant information for certain contaminant/commodity combinations for which the year-to-year variation can be significant, and the year of sampling can be some years after the year of production. It was suggested to make this field mandatory but as the year-to-year variation is not an issue for all contaminant/commodity combination and this information might not be available for the data submitter it was agreed to keep the field optional. On the suggestion to replace "year of production" by "year of harvest", it was mentioned that this is not suitable for e.g., food of animal origin. Therefore, it was concluded to propose the field name "Year of production/harvest".
- i) New field "Country of origin/production": it was stressed that this was relevant information as the country of origin might in many cases differ from the submitting country (field G) and for contaminants with regional differences, the information on country of origin/production is very relevant.
- j) New field "ML in Sampling Country/Region": it is unclear why this new field would not be possible while this information is highly relevant when considering data from different regions for ML development. An existing ML in a country might result in a bias of the results for imported products as there might be a selection of the products exported to that country to ensure compliance and therefore not reflecting the

level of contamination of the whole domestic production. The comment was made that this information on applicable MLs is available elsewhere and does therefore not need to be included in the GEMS/Food database and this could also furthermore lead to errors. It was replied that this is not straightforward as the applicable ML can evolve over time and it is relevant to know the applicable ML at the time of sampling and data reporting. A currently applicable ML cannot be as such used for evaluating all submitted data from many years, in which the ML might have changed. As this information is known to the data submitter, is relevant and is considered a minor additional burden.

- k) New field "Product type": it is clear that this field is relevant for specific commodity/contaminant combinations but not for all. In case it relates a commodity/contaminant combination for which the distinction "for further processing" or "ready to eat" is not relevant, then in the drop-down menu "not applicable" has to be chosen. In case it relates a commodity/contaminant combination for which the distinction "destined for further processing" and "ready-to-eat" is relevant but unknown then in the drop-down menu "unknown" has to be chosen. This can be clarified in a flag.
  - l) New field "Sampling Location in Production Chain": On the request to have this field optional, it is considered that this field should remain mandatory but with the option in drop-down "unknown" in case the sampling location is not known. Following the comment made as regards the option "bulk transport", it could be considered to replace it by "large bulk lots in trade/transport". Instead of the initial proposed 8 options the number of choices in the drop-down menu could be limited to 5: production site, large bulk lots in trade/transport, border (import/export), market/retail, unknown.
  - m) New field "Method of Analysis": it is agreed to provide a list of methods of analysis principles/approaches to be provided in a drop-down menu.
16. Due to time constraints, no discussion on the topics for the sections "Data selection/clean-up" and "statistical data analysis" could be discussed (in particular the merging/combining of certain topics from the section "statistical data analysis" with topics of the section "data selection/clean-up" (provided for information in Annex II), the check of the completeness of the topics for which further discussion is needed (provided for information in Annex III) and no confirmation of the correctness of the integration of the outcome of the discussions at CCCF16 into the sections "Data selection/clean up" and "statistical data-analysis"(provided for information in Annex IV),

#### **RECOMMENDATIONS TO CCCF17**

17. The following recommendations are put forward to CCCF17 for consideration and agreement:
- (i) to agree on the proposed feedback to be provided to the GEMS/Food database administrator on his preliminary feedback.
  - (ii) to agree on the proposed new work procedure as presented to the Plenary.

## ANNEX I

## Part A: Modifications to existing fields – fields with a grey background are fields where no changes are proposed.

Col	Field	Field type/ Drop-down items	Mandatory or Optional	Flag Language	Requested new language	Rationale	GEMS feedback (preliminary)
E	Local Food Identifier	Free text	Mandatory		Add flag on <i>Worksheet 2: Food Mapping</i> : "Provide a detailed name in the Local Food Identifier such as "Orange roughly" instead of "Fish."	Note: This is intended to prompt users to enter names that will be more useful for sorting and analysis.	OK
F	Serial no of the Record	Free text	Mandatory		Add flag: "One serial number (sample ID) is used for each sample. Data on different contaminants in the same sample should have the same serial number."	Provides clarity on serial no of the record.	OK
G	Country/Region	Menu	Optional		Change field name in "Submitting Country/Region" and/or Add flag: "Reflects countries or regions submitting data; this is not the country of production."	Provides clarity to submitters.	Requested flag: OK
H	Contaminant	Menu	Optional	Current flag language: "Please select a contaminant from the list . . . This is optional if a contaminant is provided on the first page."	Modified flag: "Please select a contaminant from the list. A contaminant is required, but manual entry in Column H: Contaminant is optional if a contaminant has been added on Worksheet 1: Start."	The request is to clarify language in the flag as there were questions about why a contaminant is optional.	OK
I	Food Origin	Menu: • Domestic • Imported • Mixed origin • Unknown	Optional				
J	Sampling Date	Free text (YYYY)	Mandatory				
K	Sample representativeness/reliability	Menu • Random sampling • Targeted sampling • Unknown	Mandatory		<i>Change field title:</i> Sample representativeness  <i>Change dropdown menu:</i> - no change to the dropdown menu • Random sampling • Targeted sampling • Unknown	Note: The request is to remove "reliability" from the field name and to add (routine) after random in the dropdown menu field. <i>(Proposed clarification: The term "random sampling" should be chosen for routine sampling, even if targeted at specific food types or specific importing countries. Testing a wide range of imported samples of</i>	Although this is OK in principle, the feasibility of changing a field name needs to be confirmed by developers.

					Add a flag clarifying “random sampling” and “target sampling” and provide the clarification in the instructions for electronic submission – refer to definitions of the terms in the glossary.	<i>a certain food category for the presence of a certain contaminant would be “random”. The term “targeted sampling” should be chosen for follow-up sampling following specific findings of contamination. For example, if a country identifies a sample from a particular manufacturer as having high levels of a contaminant, additional sampling of the same lot or lots produced at the same time by the same manufacturer would be “targeted”.)</i>	
L	Laboratory Identification	Free text	Optional				
M	Analytical Quality Assurance	Menu <ul style="list-style-type: none"> <li>Internal QA only</li> <li>Successful proficiency testing</li> <li>Officially accredited</li> </ul>	Optional				
N	Measurement units for Contaminant Levels	Drop-down <ul style="list-style-type: none"> <li>mg</li> <li>ug</li> <li>ng</li> <li>pg</li> <li>bg</li> </ul>	Mandatory		<ul style="list-style-type: none"> <li>mg/kg</li> <li>µg/kg</li> <li>ng/kg</li> <li>pg/kg</li> <li>Bq/kg</li> </ul>	This field is already mandatory and currently complete units are shown in the flag. The request is for complete units (mg/kg vs mg) also to appear in the rows.	Cosmetic change (the data as extracted already appear as mg/kg etc...). Will check feasibility, but OK in principle.
O	LOD	Free text	Mandatory for results not quantified if LOQ is not provided		Optional <b>Change the order of the fields: field O to come after field P</b>	Note: This can become Optional only if the LOQ is mandatory.	The order change does not bring any value added.  This change might be feasible, however the fate of legacy data with no LOQ data needs to be checked.
P	LOQ	Free text	Mandatory for results not quantified if LOD is not provided		Mandatory <b>Change the order of the fields: field P to come before field O</b>	“Mandatory” would replace “Mandatory for results not quantified if LOD is not provided”. <b>(mandatory would only apply to new submissions in order to maintain the validity of previously submitted data, without reporting of LOQ)</b>	This change might be feasible, however the fate of legacy data with no LOQ data needs to be checked.
Q	Results based on	Drop-down menu	Mandatory		<i>Change dropdown menu to:</i> <ul style="list-style-type: none"> <li>As is (raw, fresh, as</li> </ul>	Note: The request is to make changes to the drop-down menu.	<b>2 changes requested:</b>

		<ul style="list-style-type: none"> <li>•Fat content</li> <li>•Dry weight</li> <li>•As is (raw, fresh)</li> <li>•As consumed</li> </ul>			<ul style="list-style-type: none"> <li>• sold)</li> <li>• As consumed</li> <li>• Fat content <ul style="list-style-type: none"> <li>---- Fat content % [free text, allow specific # or range] to consider this information (%) in new field "compositional information"</li> </ul> </li> <li>• Dry weight <ul style="list-style-type: none"> <li>---- Water content % [free text, allow specific # or range] to consider this information (%) in new field "compositional information"</li> </ul> </li> </ul>		<p><b>First change:</b></p> <p>As is (raw, fresh) =&gt; As is (raw, fresh, as sold)</p> <p>Feasibility of change needs to be checked (with legacy data in mind).</p> <p>If not feasible, this could be addressed through a new flag.</p> <p><b>Second change:</b></p> <p>Fat /water content. Preferred option: new field compositional information will be considered.</p>
R	Portion Analyzed	<p>Menu</p> <ul style="list-style-type: none"> <li>•Edible only</li> <li>•Total food + (edible inedible)</li> </ul>	Mandatory		<p><i>Change dropdown menu to:</i></p> <ul style="list-style-type: none"> <li>•Edible only</li> <li>•Whole food (edible + inedible)</li> </ul> <p><i>Add to flag:</i></p> <p>Example: shelled nut (edible) versus unshelled nut (whole food)</p>	<p>This field already exists and is already mandatory. The request is to add examples in the flag like "shelled versus unshelled/peeled versus unpeeled" and to change Total to Whole.</p>	<p>Total to whole: cosmetic change. Will check feasibility, but OK in principle.</p> <p>Requested flag: OK</p>
S	State of food Analyzed	<p>Menu</p> <ul style="list-style-type: none"> <li>•Cooked</li> <li>•Raw</li> <li>•Unknown</li> </ul>	Optional		<p><i>Change title to:</i></p> <p>State of food analyzed (Cooked/Raw)</p>	<p>The request is to clarify that this field applies to, e.g., cooked fish versus raw fish.</p>	<p>Cosmetic change with little value added given the 1) drop-down menu clarity, and 2) loss in concision.</p> <p>Will reluctantly consider after checking feasibility.</p>
T	Results	Free text	Mandatory	<p>Current flag: Result is mandatory if LOD and LOQ are not provided.</p>	<p><i>Change flag to:</i></p> <p>"Numeric result is mandatory if LOD or LOQ are not provided."</p>	<p>For clarification. In relation to the proposed change of Field O in "mandatory". The flag although not relevant anymore for new submissions if change to Field O is accepted, remains relevant for previously submitted datasets.</p>	<p>OK</p>
U	Aggregated	Menu	Optional		Proposed to make this field		Fate of legacy data needs to be checked.

	sample	<ul style="list-style-type: none"> <li>•Individual</li> <li>•Aggregated</li> </ul>			mandatory		
V	Confidentiality of Data	Menu <ul style="list-style-type: none"> <li>•Yes</li> <li>•Blank</li> </ul>	Optional		<i>Change dropdown menu to:</i> <ul style="list-style-type: none"> <li>•Yes</li> <li>•No</li> </ul>	To improve clarity; the meaning of "blank" is unclear.	<p>Already discussed. This is sensitive. The underlying assumption is that data shared by MS is in principle (by default) not confidential. It is feared that asking data submission to specify that data are not confidential can add a burden/responsibility, or seem like the are providing a license, without being sure that are entitled to do so.</p> <p>Conclusion: given the lack of practical advantage of adding the 'no' and the potential to deter MS from submitting data, I would not consider the proposed change</p>
W	Remarks/References	Free text	Optional				

## Part B: Proposed new fields.

Col	Proposed Field	Field type/ Drop-down items	Mandatory or optional	Flag language	Requested new language	Rationale	GEMS feedback (preliminary)
--	Year of Production	Free text (YYYY)	Optional		N/A – new field	Optional – may not be known	How often will this be known, and how often will this be different from the sampling date (>1 calendar year?) Value added in EWG work VS sampling date (column J)?  Seems unnecessary.
--	Compositional Information	Free text	Optional	Information from labels such as major ingredients or percent total cocoa solids in chocolate  See field Q: add fat content or water content, as appropriate	N/A – new field	Optional --does not apply to all samples.	Will check feasibility, but OK in principle.
--	Country/Region of Origin/production	Menu <ul style="list-style-type: none"> <li>Unknown</li> <li>Countries (A-Z)</li> </ul>	Optional	Name of country of origin or production  for finished products, refer to country of origin as mentioned on the label	N/A – new field	Information may not be available	Sensitive. Will check feasibility and would consider to keep this field display only for ‘superusers’.
--	ML in Sampling Country/Region	Menu: <ul style="list-style-type: none"> <li>Yes</li> <li>No</li> <li>Unknown</li> </ul>	Mandatory	A numerical value or link to regulation can be added optionally in Remarks	N/A – new field	The submitter can be responsible for knowing whether there are MLs in the sampling country. This information will inform the EWG on whether national or regional regulations have affected contaminant levels.	No.
--	Product Type	Menu: <ul style="list-style-type: none"> <li>Destined for further processing</li> <li>Ready to eat</li> <li>Not applicable</li> <li>Unknown</li> </ul>	Optional	DFP and RTE are defined in CODEX STAN 193-1995.	N/A – new field	Optional because this does not apply to most samples.	Will check feasibility, but OK in principle.  Please elaborate what ‘Not applicable’ means in this context, noting that the field is optional, and that unknown is an option. Kindly illustrate that with an example.
--	Sampling Location in Production Chain	Menu: <ul style="list-style-type: none"> <li>Unknown</li> <li>Farm</li> <li>Bulk</li> </ul>	Mandatory		N/A – new field	The field can be mandatory with the options of Unknown and Other field	Could be considered as an <b>optional</b> field.  Too many options in my view, and could be limited

		<ul style="list-style-type: none"> <li>transport</li> <li>• Import collection</li> <li>• Industry</li> <li>• Wholesale</li> <li>• Retail</li> <li>• Other</li> </ul>					<p>to:</p> <ul style="list-style-type: none"> <li>• Production site</li> <li>• Market/retail</li> <li>• Border (import/export)</li> </ul>
--	Method of Analysis	<p>Menu</p> <ul style="list-style-type: none"> <li>• Method A</li> <li>• Method B</li> <li>• Method Z</li> <li>• Other</li> <li>• Unknown</li> </ul>	Optional		N/A – new field	<p>May provide valuable information in conjunction with LOQ/LOD. The dropdown menu should provide options between methods of analysis principles/approaches and not provide a very long list of methods, specifying all possible variants of a certain method of analysis principle/approach</p>	<p>Will check feasibility, but OK in principle.</p> <p>The EWG is requested to come up with a comprehensive list of methods for consideration. Suggestion is to group these to avoid a too extensive list of options (ex: mass spectrometry applications, rapid tests, other methods)</p>

## ANNEX II

### Topics for the sections “Data selection/clean-up” and “statistical data analysis” (in particular the merging/combining of certain topics from the section “statistical data analysis” with topics of the section “data selection/clean-up”)

#### Data Selection /clean-up of data

General considerations

Selection and clean-up - Handling of data

- with a lack of information
- for which an error in reporting is assumed
- originating from suspected fraudulent/ adulterated samples
- from targeted sampling
- outliers/extreme values (including methods to determine outliers/extreme values)
- Limit of Quantification (LOQ) and Limit of Detection (LOD) considerations

#### Data analysis: generating overview of data

- Overview which countries, how many data points, which years, period of data coverage
- Decision on geographical coverage of the provided occurrence data (including consideration of combining or keeping separate different datasets)
- Decision on period coverage of the provided occurrence data

#### Statistical analysis of occurrence data / handling of datasets for ML development

General considerations

Sufficient number of samples

- minimum number of samples for estimating high percentile values

Handling of datasets

- with low number of data points
- with data on individual food(s) are insufficient, but data for the food group are sufficient
- with a large proportion of left-censored data (including use of substitution methods)

Conducting statistical analysis

- Drawing charts/graphs and plots on distribution of occurrence data

Data aggregation and calculation of descriptive statistics

Calculation of rejection rates at hypothetical MLs

- Estimation of hypothetical MLs
- Calculation of rejection rates at the hypothetical MLs
- Assessment of impact of an ML on rejection rate
- Improvement of calculation of rejection rates

Calculation of effects of MLs on the reduction of dietary exposure at hypothetical MLs

- Calculation of dietary exposure and reduction at hypothetical MLs
- Assessment of impact of ML on dietary exposure
- Improvement of calculation of exposure reduction rates

#### Data presentation in EWG reports to CCCF

#### ANNEX: Glossary of terms

#### Merging /combining of certain topics from the section “data analysis” with topics of the section “data selection/clean-up”:

- Move the part on “Determination of outliers/extreme values” to “data selection and clean-up” (as certain aspects are related to aspects of data clean-up – outliers due to e.g., error in reporting, fraudulent /adulterated samples).
- Combine the part on “Handling of multiple datasets” with “generating overview of data” (geographical coverage, period coverage).

**ANNEX III****Topics for which further discussion is needed.**

- Minimum number of samples for estimating high percentile values with high confidence (REP23/CF16, paras 93 and 94, para 98 (vi) (a)).
- Further guidance to be provided on which dataset the ML should be based or to which database should be given priority for ML development (combined dataset, dataset showing the higher contamination patterns as long as the commodity was produced through good practice, datasets from major producing countries or regions, datasets from importing countries reflecting the levels of a contaminant in a commodity in international trade, dataset to be used to be decided on a case-by-case ) (REP23/CF16, para 98 (vi) (b)).
- Further consider the role of the Committee in calculating dietary exposure reduction rates when considering MLs. (calculation of dietary exposure is a risk assessment function that should be undertaken by JECFA and JECFA provides the scientific advice on which the risk management decisions of the Committee are based – it is important to clarify the roles of JECFA and CCCF as risk assessor and risk manager respectively, in the calculation of dietary exposure reduction rates when considering MLs (REP23/CF16, paras 90 and 91 and para 98 (vi) (c)).
- More structured process for elaborating calls for data (REP23/CF16, para 98 (viii)).
- Identification of appropriate rejection rates in ML establishment (guidance on elements which should be considered to define the appropriate rejection rate) (CX/CF 22/15/14 chapter IV of appendix I, CF16/CRD06 para 32).
- Appropriateness of GEMS/Food market-based cluster diets for ML elaboration (reconcile realistic estimates from national consumption data with the “supply utilization market data in GEMS/Food cluster diets (e.g., sugar, spices/herbs, teas, coffees).

**ANNEX IV****Integration of the outcome of the discussions at CCCF16 into the sections “Data selection/clean up” and “statistical data-analysis”.**

- **Preamble:** Development of a discussion paper that explores preliminary occurrence data, exposure data, and global significance of the contaminant-commodity/food combination. **The discussion paper needs also to consider the data availability and quality to enable an informed decision on possible new work (REP23/CF16 para 95 and para 98 (ix))**
- **Statistical analysis – minimum number of samples for estimating high percentile values**  
CCCF16 agreed to a **provisional** minimum number of 59 samples for a 95th percentile estimation with 95% confidence (option 1 in table 1) (REP23/CF16 paragraphs 93, 94, 98 (vi) (a)). However, CCCF16 also agreed that the minimum number of samples for estimating high percentile values with high confidence needs to be further discussed (REP23/CF16 paragraphs 93 and 98 (vi) (a))
- **Statistical analysis – Handling of multiple datasets**  
CCCF16 concluded that at this stage the combined global dataset is to be used for the development of the ML and the individual datasets per year or per region are provided for additional consideration in the ML development.  
  
But at this stage there would be no guidance given on which dataset the ML development should be based or to which database should be given priority for ML development (combined dataset, dataset showing the higher contamination patterns as long as the commodity was produced through good practice, datasets from major producing countries or regions, datasets from importing countries reflecting the levels of a contaminant in a commodity in international trade, dataset to be used to be decided on a case-by-case).  
  
This topic requires further discussion after CCCF17 (REP23/CF16, para 98 (vi) (b))
- **Statistical analysis – Calculation of effects of MLs on the reduction of dietary exposure at hypothetical MLs**  
Further consideration of the role of the Committee in calculating dietary exposure reduction rates when considering MLs is needed. (calculation of dietary exposure is a risk assessment function that should be undertaken by JECFA and JECFA provides the scientific advice on which the risk management decisions of the Committee are based – it is important to clarify the roles of JECFA and CCCF as risk assessor and risk manager respectively, in the calculation of dietary exposure reduction rates when considering MLs (REP23/CF16, paras 90 and 91 and para 98 (vi) (c))