

CODEX ALIMENTARIUS COMMISSION



Food and Agriculture
Organization of the
United Nations



World Health
Organization

E

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

CL 2021/78-CF

October 2021

TO: Codex Contact Points
Contact Points of international organizations having observer status with Codex

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

SUBJECT: **Request for comments on the guidance on data analysis for development of maximum levels and for improved data collection**

DEADLINE: **15 December 2021**

BACKGROUND

1. For background information, please refer to the discussion held at the 14th Session of the Codex Committee on Contaminants in Foods (REP21/CF, paras. 186-210) based on the information presented in working paper CX/CF 21/14/15 available on the CCCF14 webpage¹.

REQUEST FOR COMMENTS

2. Codex members and observers are invited to provide comments on the different points identified in relation to the guidance on data analysis for development of MLs and improved data collection for contaminants in food and feed as follows:
 - A. Criteria for the establishment of maximum levels in food and feed
 - B. Improved data collection
 - C. Handling/acceptance of data within a dataset
 - D. Important topics to be considered for data analysis
 - E. Guidance on how to present the data in reports to electronic working groups to the Codex Committee on Contaminants in Foods
 - F. Issues identified in the data analysis for possible maximum levels for lead and total aflatoxins
3. For convenience, the Annex to CX/CF 21/14/15 is presented in the Annex to this CL.
4. Comments submitted will be considered by the Electronic Working Group established by CCCF14 to prepare guidance on data analysis for development of MLs and for improved data collection based on the comments provided at CCCF14 and those in reply to this circular letter.

GUIDANCE ON THE PROVISION OF COMMENTS

5. Comments should be submitted through the Codex Contact Points of Codex members and observers using the OCS.
6. Contact Points of Codex members and observers may login to the OCS and access the document open for comments by selecting "Enter" in the "My reviews" page, available after login to the system.
7. Contact Points of Codex members and observers organizations are requested to provide proposed changes and relevant comments/justifications on a specific paragraph (under the categories: editorial, substantive, technical and translation) and/or at the document level (general comments or summary comments). Additional guidance on the OCS comment categories and types can be found in the OCS [Frequently Asked Questions \(FAQs\)](#).
8. Other OCS resources, including the user manual and short guide, can be found at the following link: <http://www.fao.org/fao-who-codexalimentarius/resources/circular-letters/en/>.
9. For questions on the OCS, please contact Codex-OCS@fao.org.

¹ <https://www.fao.org/fao-who-codexalimentarius/meetings/detail/en/?meeting=CCCF&session=14>

ANNEX
(CX/CF 21/14/15, Annex)
ORIGINAL LANGUAGE ONLY

A) CRITERIA FOR THE ESTABLISHMENT OF MAXIMUM LEVELS IN FOOD AND FEED²

Selection of criteria has been made of relevance for improved data collection and analysis of data for setting MLs

- Validated qualitative and quantitative analytical data on representative samples should be supplied. Information on the analytical and sampling methods used and on the validation of the results is desirable. A statement on the representativeness of the samples for the contamination of the product in general (e.g. on a national basis) should be added. The portion of the commodity that was analyzed and to which the contaminant content is related should be clearly stated and preferably should be equivalent to the definition of the commodity for this purpose or to existing related contaminant regulation.
- Information on appropriate sampling procedures should be supplied. Special attention to this aspect is necessary in the case of contaminants that may not be homogeneously distributed in the product (e.g. mycotoxins in some commodities).
- MLs should be set as low as reasonably achievable and at levels necessary to protect the consumer. Providing it is acceptable from the toxicological point of view, MLs should be set at a level which is (slightly) higher than the normal range of variation in levels in food and feed that are produced with current adequate technological methods, in order to avoid undue disruptions of food and feed production and trade. Where possible, MLs should be based on GMP and/or GAP considerations in which the health concerns have been incorporated as a guiding principle to achieve contaminant levels as low as reasonably achievable and necessary to protect the consumer. Foods that are evidently contaminated by local situations or processing conditions that can be avoided by reasonably achievable means shall be excluded in this evaluation, unless a higher ML can be shown to be acceptable from a public health point of view and significant economic aspects are at stake.
- Proposals for MLs in products should be based on data from various countries and sources, encompassing the main production areas/processes of those products, as far as they are engaged in international trade. When there is evidence that contamination patterns are sufficiently understood and will be comparable on a global scale, more limited data may be enough.
- MLs may be set for product groups when sufficient information is available about the contamination pattern for the whole group, or when there are other arguments that extrapolation is appropriate.
- Numerical values for MLs should preferably be regular figures in a geometric scale (0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5 etc.), unless this may pose problems in the acceptability of the MLs.

B) IMPROVED DATA COLLECTION

Important elements to be provided when reporting occurrence data

- 1) information on the stage in production and production chain where the sampling took place (farm, wholesale, import, retail) and location (country/region) of sampling. If known, origin of product sampled.
- 2) information on type of sampling : targeted sampling, suspect sampling, random sampling
- 3) food and feed to be correctly identified and reported with detailed information on the food or feed concerned (correct identification, state of the food/feed (fresh, dried, ready-to-eat, etc.)
- 4) information on the portion of food analysed (e.g. peeled or not, edible part or whole fruit, etc..)
- 5) the unit of the data provided and the basis on which the data are expressed (e.g. fat basis vs whole weight)
- 6) information on the methods of analysis (and their validation) used for generating occurrence data with information on the LOQ/LOD of the method

² Reference is made to the criteria for the establishment of maximum levels in food and feed as provided for in Annex I of CXS 193-1995 General Standard for Contaminants and Toxins in Food and Feed

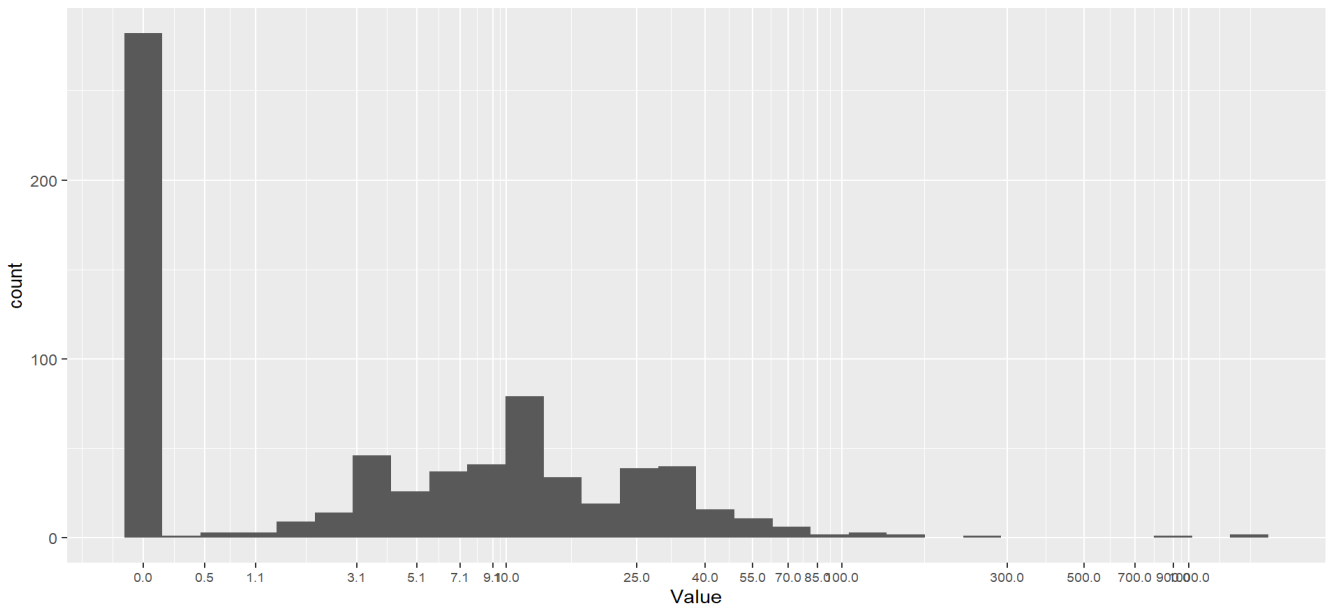
C) HANDLING/ACCEPTANCE OF DATA WITHIN A DATASET**1) Handling of outliers/extreme values**

When to consider data as outliers/extreme values?

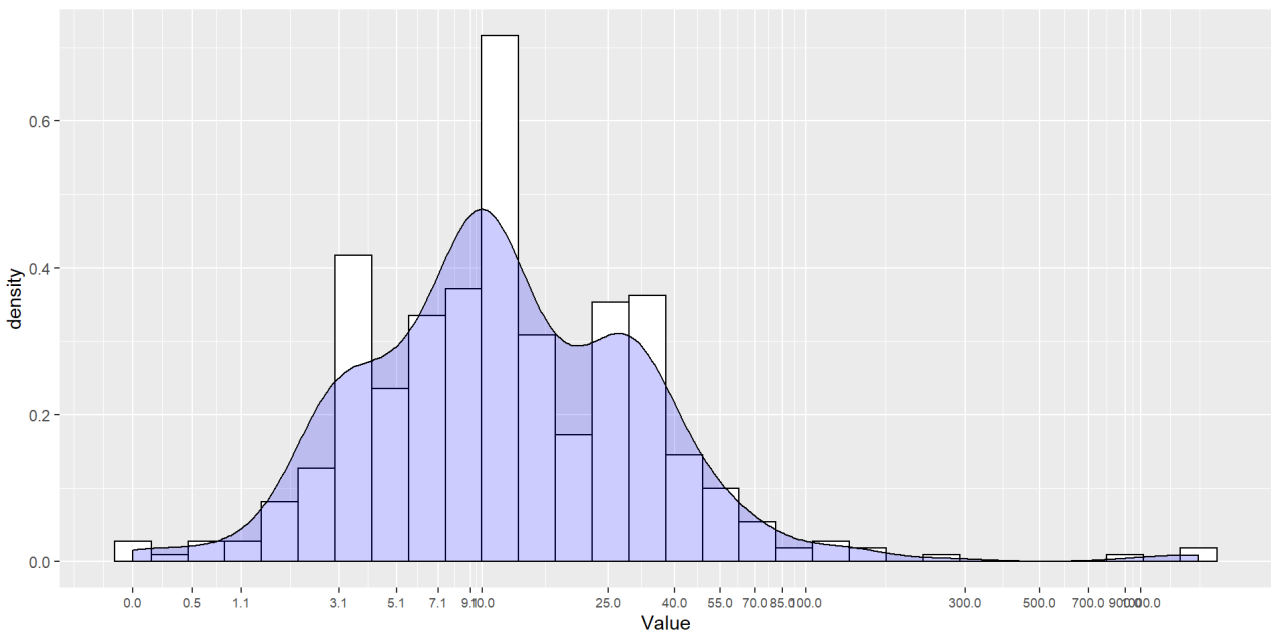
For consideration: In case the data are outside the range of distribution of the data and no justification can be provided for these extreme results (such as data from a year with extreme weather conditions, data from a specific region/continent, ...)

Example EU data on sum of T-2 and HT-2 toxin in oat milling products (717 results of which 438 quantified results)

Histogram: all results



Histogram/density of quantified results



In case no justification for the data with levels $> 500 \mu\text{g}/\text{kg}$ can be provided these data could be considered as outliers.

2) Handling of data for which it can be reasonably assumed that the unit of the data provided or the basis on which the data are reported (e.g. fat basis vs whole weight) is not correct.

If there are clear indications that the unit in which the data are expressed is incorrect or the basis on which the data are expressed is incorrect, these data should be excluded from further data analysis.

Examples of “clear indications”

* Levels within a data set of 200 results are in the range of 0 to 20. All data are expressed as $\mu\text{g}/\text{kg}$, except 5 quantified data points expressed as mg/kg . When putting these data in a frequency distribution curve (see a) they would be identified as possible outlier

* Levels from a food with a typical fat content of 5 % within a data set of 200 results of which all data are expressed on whole weight. 195 results are falling in the range of 0-20 mg/kg , however 5 data points are falling within in the range of 100–400 mg/kg . When putting these data in a frequency distribution curve (see a) they would be identified as possible outlier

3) Lack of information on data provided

It has to be considered to which extend the missing information makes the data unusable.

Examples of missing information by which data cannot be used for further data analysis:

- All data from a dataset are reported as < LOQ and the LOQ is not provided (more information in point 2 in Chapter D)
- the unit in which the result is reported is missing or the basis on which the result is expressed
- the state of the food sampled (dried fresh)

Examples of missing information but the data could still be used for further data analysis:

- sampling information: type of sampling, year of sampling, location of sampling, ...
- method of analysis used

4) Handing of the data not provided to the GEMS/food

It has to be considered if these data can be taken into further data analysis

- in case there are only limited data available in the GEMS/food database, it could be considered useful to use these data in further data analysis.
- in case there are extensive data available in the GEMS/food database, it could be considered not to use these data in further data analysis (and certainly not in case the data do not show a contamination pattern different than the data available in the GEMS/food database).

5) Handling of datasets with a different contamination pattern (e.g. as consequence of originating from different regions, different production years)

Guidance should be provided on when to combine or keep separate such datasets for assessment

- if datasets from different regions/continent in the world show a different contamination pattern and a valid reasoning for the difference can be provided (e.g. different climate conditions, different production conditions/techniques) , then the datasets could be kept separate for assessment.

D) IMPORTANT TOPICS TO BE CONSIDERED FOR DATA ANALYSIS

1) Minimum number of samples needed for the use of percentiles

Background information

In order to apply the above criterion “*MLs should be set at a level which is (slightly) higher than the normal range of variation in levels in food and feed*”, high percentiles are used to define that level. The reliability of high percentiles is related to the number of data used to calculate them. Percentiles calculated on a number of subjects should be treated with caution as the results may not be statistically robust.

A clear indication concerning the minimum number of observations necessary to estimate a given percentile is not provided in literature. Different options can be used, none of them being a widely accepted standard.

A very simple option is to require that the calculated percentile must at least be different from the maximum value within the sample. This means that at least 20 observations are needed to identify the single observation at the 95th percentile and 100 observations are needed for the 99th percentile.

In statistics, the coverage probability of a confidence interval is the probability that the interval contains the true value of interest (e.g. 95th or 99th percentiles). When the number of observations is not large enough, the coverage probability may not attain the nominal value, and drops below, for example, 95%. This is more likely to occur at high percentiles, e.g. 95th or 99th. Therefore, the coverage probability has been used to set guidelines to determine the minimum number of samples for which (extreme) percentiles can be computed. In the case of significance level (α) being set at 0.05 to determine a 95% confidence interval, the coverage probability should target 95%. In this case, this is achieved for $n \geq 59$ and $n \geq 298$ for the 95th or 99th percentiles, respectively.

2) Limit of Quantification (LOQ) considerations

Several situations applicable to datasets provided can occur and the guidelines to be elaborated should provide guidance on how to handle the datasets in the different situations

- No LOQ provided
 - Dataset contains (nearly) all quantified results
 - Dataset contains a significant part of left-censored data (i.e. < LOQ) and no LOQ provided

In the above situations where the LOQ is not provided, should the guidance provide for different conclusions as regards how to handle the dataset in case the quantified results (significantly lower than the ML under consideration) in the dataset provide an indication that the LOQ is (very) low compared to datasets where the quantified results do not provide that indication.

- LOQ provided
 - Dataset with LOQ significantly lower than the ML under consideration
 - Dataset with LOQ in the range of the ML under consideration
 - Dataset with LOQ above the ML under consideration

In the above situations where the LOQ is provided, should there be guidance on cut-offs to be used for the LOQ on the analytical results dataset used for the ML development?

Should the guidance provide for different conclusions as regards how to handle the dataset in case the dataset contains nearly all quantified results compared to a dataset with nearly all left-censored data?

3) Using data sets with a large proportion of left-censored data for ML development

In certain cases, the analytical results for one specific contaminant are produced with a battery of different analytical methods and/or the same analytical method but with very different sensitivities. As a consequence, there could be a wide range of limits of detection (LOD) and limits of quantification (LOQ) for a particular contaminant and food matrix in a given dataset, composed of datasets from different sources. This situation is particularly relevant when the occurrence datasets used for the ML development contain a high number of non-quantified/non-detected data (left-censored data).

The standard approach to deal with left-censored data is the use of the substitution. In this method, at the lower-bound (LB), results below the LOQ and LOD are replaced by zero; at the upper-bound (UB) the results below the LOD are replaced by the numerical value of the LOD and those below the LOQ are replaced by the value reported as LOQ. Additionally, as a point estimate between the two extremes, the middle-bound (MB) scenario is calculated by assigning a value of $LOD/2$ or $LOQ/2$ to the left-censored data.

4) Geographical coverage of the provided occurrence data

Guidance should be provided to evaluate the appropriateness of the geographical coverage of the provided data for ML development and a procedure should be developed for situations for which it is concluded that the available data do not provide a sufficient/appropriate geographical coverage.

5) Period coverage of the provided occurrence data

Guidance should be provided in which situation it might be required that the provided occurrence data relate to several production years for ML development (can be different for different types of contaminants: mycotoxins, plant toxins, processing contaminants, environmental contaminants in function of the assumed year to-year variation or evolution of contamination in time)

6) Data sets with low number of data (e.g. less than 60) for development of ML

Guidance could be given in which situations it can be concluded that the data, despite the low number, are sufficient for the development of an ML (e.g. despite limited number good geographical coverage, no large variation in occurrence observed despite data originating from different regions/from different years, etc).

E) GUIDANCE ON HOW TO PRESENT THE DATA IN EWG REPORTS TO CCCF

It is important that the data are represented in such a way in the EWG report to CCCF to enable an informed discussion on appropriate MLs to be established.

The detail of reporting depends on the amount of data available and also of the nature of the contaminant.

Elements of consideration (not-exhaustive)

- if there is a significant year-to-year variation in occurrence it is appropriate to provide an analysis of the data per year.
- if there is a significant difference in contamination pattern between regions of e.g. climate conditions or production methods, it is appropriate to provide an analysis of the data per year;
- the description of the data should provide a clear view on the data set e.g;

* Number and proportion of positive (quantified results)

* Mean, median and range of positive results

* P90, P95, P99

* histograms/density of positive results (example see

F) ISSUES IDENTIFIED IN THE DATA ANALYSIS FOR POSSIBLE MLs OF LEAD (agenda item 8, CX/CF 21/14/8) and MLs of TOTAL AFLATOXINS (agenda item 10 (a) CX/CF 21/14/10 – Part I) NOT MENTIONED BEFORE**1) Application of different rejection rates for different types of products and contaminants, deviating from the usual rejection rate of 5%**

At the 13th session of CCF it was clarified that the basis on which the MLs should be proposed (i.e. rejection rate, occurrence data and reduction risk) was outside the scope of the guidance (§ 162, REP19/CF)

However there is the explicit request to the CCCF in relation with the discussion on MLs for lead and total aflatoxins whether different rejection rates should be applied for different types of products and contaminants. Therefore, CCCF might agree that it is appropriate to provide in this guidance, elements which should be taken into account to define the appropriate rejection rate. This should increase the transparency on the basis on which grounds a maximum level has been set.

Possible elements for consideration (not exhaustive)

- nature of the product:
 - o raw cereals of which already large part is used for feed: non-compliance with the food ML might not necessarily result in economic damage as it can still be used as feed.
 - o processed products intended for human consumption: non-compliance with the food ML will result in economic damage as possible alternative uses will result in lower return or in certain cases the lot has to be destroyed.
- different regional contamination patterns:
 - o worldwide dataset might have a rejection rate lower than 5 % at a certain ML while regional datasets might have for the same ML much different (lower or higher) rejection rate.