



PROGRAMME MIXTE FAO/OMS SUR LES NORMES ALIMENTAIRES

COMITÉ DU CODEX SUR LES CONTAMINANTS DANS LES ALIMENTS

Dix-huitième session

23-27 juin 2025

Bangkok (Thaïlande)

ORIENTATION SUR L'ANALYSE DES DONNÉES EN VUE DE L'ÉLABORATION DE LIMITES MAXIMALES ET POUR L'AMÉLIORATION DE LA COLLECTE DES DONNÉES

(Préparée par le groupe de travail électronique présidé par l'Union européenne
et co-présidé par le Japon, les Pays-Bas et les États-Unis d'Amérique)

GÉNÉRALITÉS

1. Lors de la 12^e session¹, le CCCF a examiné la proposition du Secrétariat du JECFA de rédiger une orientation générale pour l'analyse des données en vue de l'élaboration de limites maximales (LM), car il a été observé que les GTE adoptaient différentes approches. Ces différences se rapportaient, par exemple, au traitement des données d'occurrence sans information sur la limite de quantification (LOQ). Une orientation générale aiderait les futurs GTE à adopter des approches cohérentes en matière d'analyse des données. Le CCCF a convenu d'établir un GTE présidé par l'UE, coprésidé par les États-Unis d'Amérique, les Pays-Bas et le Japon, travaillant en anglais, pour préparer un document de travail.
2. Au cours de sa 13^e session², l'UE, en sa qualité de Présidente du GTE, a informé le CCCF qu'il n'avait pas été possible de préparer à temps un document de travail pour examen par le GTE établi. Par conséquent, un document préparé par l'UE en sa qualité de Présidente du GTE et contenant une liste non exhaustive de sujets qui pouvaient être considérés comme étant couverts par les directives générales pour l'analyse des données en vue de l'élaboration des LM a été présenté, et le CCCF a convenu d'étendre la portée des travaux en vue de traiter également l'amélioration de la collecte des données.
3. Lors de la 14^e session³, le CCCF a convenu que les travaux devraient porter en priorité sur la collecte, l'analyse et la présentation des données, et que la discussion sur les éléments à prendre en considération, tels que les taux de rejet appropriés, ne serait pas poursuivie. La lettre circulaire CL 2021/78 CF⁴, accompagnée de l'Annexe au document CX/CF 21/14/15 en annexe, a été diffusée en octobre 2021, assortie d'une demande d'observations sur l'Orientation sur l'analyse des données en vue de l'élaboration de limites maximales et pour l'amélioration de la collecte des données.
4. La 15^e session du CCCF⁵ a convenu du statut, des buts/objectifs et des utilisateurs cibles devant être exposés dans le Préambule, et de la structure et du contenu du document d'orientation, étant entendu qu'un réajustement ultérieur pourrait s'avérer nécessaire à la suite de la discussion au sein du GTE.
5. La 16^e session du CCCF⁶ a convenu des changements à apporter au gabarit de la base de données GEMS/Aliments, présentés sous forme de recommandations à l'administrateur du GEMS/Aliments pour révision. Une fois que l'administrateur de la base de données GEMS/Aliments aura indiqué les recommandations susceptibles d'être mises en œuvre de manière efficace et le calendrier de cette mise en œuvre, la section «Collecte, soumission et extraction des données» devra être mise à jour.
6. Toutefois, l'administrateur de GEMS/Aliments a été consulté beaucoup trop tard par le président du GTE et, par conséquent, la section «Collecte et soumission des données» n'a pas pu être mise à jour en tenant compte de ses

¹ REP18/CF, par. 155-156

² REP19/CF, par. 156-165

³ REP21/CF, par. 186-210

⁴ <https://www.fao.org/fao-who-codexalimentarius/resources/circular-letters/en/>

⁵ REP22/CF15, par. 202-208

⁶ REP23/CF16 par. 98 (i) et (ii)

observations ni communiquée au GTE pour observations ni soumise à la 17^e session du CCCF en vue de sa finalisation.

7. La 17^e session du CCCF⁷ a convenu du retour d'information à fournir à l'administrateur de la base de données GEMS/Aliments à partir des observations préliminaires sur les changements proposés par la 16^e session du CCCF sur le gabarit de la base de données GEMS/Aliments et a approuvé la nouvelle procédure des travaux proposée pour finaliser l'orientation sur l'analyse des données pour l'élaboration des LM et l'amélioration de la collecte des données.
8. Étant donné que l'Orientation sur l'analyse des données pour l'élaboration des LM et l'amélioration de la collecte des données sont en cours d'élaboration pour la procédure de travaux internes du CCCF, ces travaux pourraient être entrepris dans le cadre d'un groupe de travail pré-session (qui pourrait fonctionner en mode physique ou virtuel) ou d'un groupe de travail intersession (ci-après dénommé «le GT»). Il est envisagé de convoquer des réunions de GTV pour parler des travaux sur certaines sections du document d'Orientation et les faire avancer. Les résultats des GTV seront diffusés pour observations par Lettre circulaire, et ces observations seront examinées par un GT pré-session (virtuel ou physique) ou un GT intersession. La plateforme du GTE du Codex sera également utilisée pour faciliter ces travaux.
9. Lors de la 17^e session du CCCF⁸, il a été discuté et convenu que les travaux préalables à la 18^e session du CCCF seraient axés sur:
 - (i) la finalisation des modifications apportées au gabarit de la base de données GEMS/Aliments et à l'orientation correspondante. Il est prévu d'organiser une réunion du groupe de travail virtuel (GTV) pour parler de la collecte et de la soumission des données, ainsi que de leur extraction; et
 - (ii) la discussion sur la structure et le contenu du document principal pour la section «Sélection et nettoyage des données et analyse statistique des données», avec une décision sur les questions les plus complexes qui seront traitées à l'avenir dans une ou plusieurs annexes distinctes du document principal. Il est prévu d'organiser un autre GTV pour parler de cette question.
10. Après les discussions au sein du GTV, on pourrait examiner s'il serait judicieux de procéder à un cycle d'observations au sein du GTE avant que le document principal sur ces questions ne soit finalisé pour être diffusé sous forme de Lettre circulaire en vue de recueillir des observations.
11. La 17^e session du CCCF⁹ a convenu à l'intention d'élaborer un document qui fournirait une orientation pratique au GTE chargé de l'analyse des données en vue de l'élaboration des LM, pour examen par la 18^e session du CCCF. La 17^e session du CCCF a également convenu qu'à l'avenir, les questions plus complexes identifiées seront traitées dans des annexes distinctes qui seront élaborées et débattues après la 18^e session du CCCF, une fois que l'Orientation principale aura été définie, ainsi que les questions identifiées en vue d'une discussion future.
12. Une première réunion du GTV a eu lieu sur la partie «collecte, soumission et extraction des données» de l'Orientation sur l'analyse des données pour l'élaboration des limites maximales et pour l'amélioration de la collecte des données le 31 octobre 2024, avec les États-Unis en tant que président du GTV. Les observations formulées par les participants à la réunion du GTV et les observations supplémentaires postées sur la plate-forme du Forum du Codex par le Japon ont été pris en compte dans la version de l'Orientation disponible à l'Appendice I.
13. Une deuxième réunion du GTV a eu lieu sur les parties combinées «sélection des données/nettoyage des données» et «analyse statistique» de l'Orientation sur l'analyse des données pour l'élaboration des limites maximales et pour l'amélioration de la collecte des données le 26 février 2025, avec les Pays-Bas et le Japon en tant que présidents du GTV. Les observations formulées par les participants à la réunion du GTV et les observations supplémentaires postées sur la plate-forme du Forum du Codex par la Thaïlande, les États-Unis, Singapour, le Brésil et l'International Council of Beverages Associations (ICBA) ont été prises en compte dans la version de l'Orientation disponible à l'Appendice I.
14. Les champs actuels pour les résultats des analyses individuelles dans le gabarit de la base de données GEMS/Aliments sont énumérés dans le Tableau d'orientation 1 du document d'Orientation de l'Appendice I. L'administrateur de la base de données GEMS/Aliments a informé le Président du groupe de travail que les changements convenus pour le gabarit de la base de données GEMS/Aliments ne pourraient pas être mis en œuvre avant la 18^e session du CCCF. Par conséquent, le texte entre crochets (simples ou doubles) indique les changements qui ont été convenus, mais qui n'ont pas encore été mis en œuvre. Le texte bleu entre crochets simples indique les modifications basées sur les changements qui pourraient être mis en œuvre dans une première phase (le calendrier sera communiqué à la 18^e session du CCCF). Le texte violet entre doubles crochets indique les modifications qui pourraient être apportées ultérieurement.

POINTS DE DISCUSSION

⁷ REP24/CF17, par. 140, 141 et 145

⁸ REP24/CF17, par. 142

⁹ REP24/CF17, par. 144

15. Discussion sur les observations reçues avant et pendant la 18^e session du CCCF sur la partie principale de l’Orientation sur l’analyse des données pour l’élaboration des LM et pour l’amélioration de la collecte des données, telle que prévue dans l’Appendice I.
16. Discussion sur les rôles du JECFA et du CCCF en tant qu’évaluateur et gestionnaire des risques, respectivement, dans le calcul des taux de réduction de l’exposition alimentaire lors de l’examen des LM et dans quelle mesure l’évaluation préliminaire de l’exposition du produit de base cible peut également être effectuée en tant qu’information de référence dans les GTE chargés de recommander des LM si les ressources sont disponibles (paragraphe 156 et 157 et Annexe IV de l’Orientation dans l’Appendice I)
17. Discussion sur la poursuite des travaux sur les Annexes de l’Orientation, comme prévu à l’Appendice I, et sur les questions identifiées pour une discussion future, comme prévu à l’Appendice II. Dans le cadre de cette discussion, il convient d’envisager la fusion de l’Annexe III «Génération de diagrammes/graphiques et de courbes se rapportant à la distribution des données d’occurrence» avec l’Annexe V «Présentation de l’analyse des données/analyse statistique», étant donné que la visualisation de la distribution des données d’occurrence est un aspect important de la présentation des données d’occurrence.

CONCLUSION

18. Étant donné que l’Orientation a été développée pour la procédure des travaux internes du CCCF, la partie principale de l’Orientation sur l’analyse des données pour l’élaboration des LM et pour l’amélioration de la collecte des données pourrait être approuvée lors de la 18^e session du CCCF après avoir pris en compte les observations reçues avant et pendant la 18^e session du CCCF, étant entendu que le document sera mis à jour suite à la mise en œuvre des changements convenus au gabarit de base de données GEMS/Aliments et grâce à l’expérience acquise par les GTE dans l’application de l’Orientation.

RECOMMANDATIONS

19. Sur la base des questions soulevées dans les points de discussion (paragraphe 15 à 17) et de la conclusion (paragraphe 18), la 18^e session du CCCF est invitée à:
 - (i) approuver la partie principale de l’Orientation sur l’analyse des données pour l’élaboration de limites maximales et pour l’amélioration de la collecte des données prévue dans l’Appendice I;
 - (ii) clarifier les rôles du JECFA et du CCCF en tant qu’évaluateurs et gestionnaires de risques, respectivement, dans le calcul des taux de réduction de l’exposition alimentaire lors de l’examen des LM et dans quelle mesure l’évaluation préliminaire de l’exposition du produit de base cible peut également être réalisée en tant qu’information de référence dans les GTE chargés de recommander des LM si les ressources sont disponibles (voir également le 3^e tiret de l’Appendice II);
 - (iii) envisager la poursuite des travaux sur les Annexes à la partie principale de l’Orientation figurant à l’Appendice I;
 - Si des Annexes et le cas échéant, quelles Annexes sont nécessaires.
 - Si des Annexes et le cas échéant, quelles Annexes pourraient être combinées.
 - Détermination du calendrier pour la poursuite des travaux sur les Annexes.
 - (iv) envisager la poursuite des travaux sur les questions identifiées comme devant faire l’objet d’une discussion future et figurant à l’Appendice II.
 - Détermination des questions sur lesquelles il convient de poursuivre la discussion.
 - Détermination de l’ordre de priorité des questions à traiter.
 - Détermination du calendrier pour la poursuite des travaux sur ces questions.

LISTE DES APPENDICES

- APPENDICE I: Orientation concernant l’analyse des données pour l’élaboration de limites maximales et pour l’amélioration de la collecte des données
- APPENDICE II: Questions identifiées en vue d’une éventuelle discussion future
- APPENDICE III: Liste des Membres du groupe de la plate-forme du forum du Codex «CCCF18 - Orientation sur l’analyse des données pour l’élaboration de limites maximales et pour l’amélioration de la collecte des données»

APPENDICE I

ORIENTATION SUR L'ANALYSE DES DONNÉES EN VUE DE L'ÉLABORATION DE LIMITES MAXIMALES ET POUR L'AMÉLIORATION DE LA COLLECTE DES DONNÉES

TABLE DES MATIÈRES

	Paragraphes
Préambule	par. 1-7
Collecte et soumission de données	par. 8-16
Compléter le gabarit de GEMS/Aliments	par.17-52
Extraction de données	par. 53-57
Sélection/nettoyage de données	
Considérations générales	par. 58
Origine des données	par. 59-60
Nettoyage de données	par. 61-65
Traitement de données	
Informations manquantes	par. 66-73
Informations erronées	par. 74-77
Données pour lesquelles l'information sur la partie analysée n'est pas claire	par. 78
Données provenant d'échantillons soupçonnés d'être frauduleux ou économiquement falsifiés	par. 79
Données provenant d'un échantillonnage ciblé	par. 80-81
Détermination et traitement des valeurs aberrantes/extrêmes	par. 82-92
Considérations relatives à la limite de quantification (LOQ) et à la limite de détection (LOD)	
Exclusion sur la base de la LOQ	par. 93-96)
Somme des composants et LOQ	par. 97
Analyse des données: génération d'un aperçu de données	
Aperçu présentant quels pays, nombre de points de données, période de couverture des données	par. 98
Couverture géographique des données d'occurrence fournies	par. 99-104
Période couverte par les données d'occurrence fournies	par. 105-110
Analyse statistique des données d'occurrence et traitement d'ensembles de données pour l'élaboration des LM	
Considérations générales	par. 111-113
Nombre minimum de points de données pour l'estimation des valeurs de centile élevées	par. 114
Traitement des ensembles de données	
Traitement des ensembles de données avec un nombre de points de données faible	par. 115-119
Traitement des ensembles de données lorsque les données disponibles sur les produits de base/aliments individuels sont insuffisantes, mais que les données de pour le groupe d'aliments sont suffisantes	par. 120-123
Traitement des ensembles de données (y compris l'utilisation de méthodes de substitution)	

comportant une grande proportion de données censurées à gauche	par. 124-126
Méthodes de substitution pour les ensembles de données comportant une grande proportion de données censurées à gauche	par. 127-130

Décision relative au traitement d'ensembles de données multiples

Analyse des ensembles de données individuels et combinés, et prise de décision sur la nécessité de comparer les ensembles de données individuels avant de les combiner, en particulier lorsque les schémas de distribution sont significativement différents.	par. 131-136
---	--------------

Cas où l'analyse d'ensembles de données individuels est recommandée	par. 137-140
---	--------------

Réalisation d'une analyse des données par visualisation	par. 141
--	----------

Agrégation des données et calcul des statistiques descriptives	par. 142-144
---	--------------

Calcul des taux de rejet à des LM hypothétiques

Estimation des LM hypothétiques	par. 145-148
Calcul des taux de rejet à des LM hypothétiques	par. 149-150
Évaluation de l'impact d'une LM sur le taux de rejet	par. 151-152
Amélioration du calcul des taux de rejet	par. 153-155

Calcul préliminaire des effets des LM sur la réduction de l'exposition alimentaire au contaminant d'un ou de produits de base cibles, à des LM hypothétiques

Calcul de l'exposition alimentaire et de la réduction à des LM hypothétiques	par. 156
Amélioration du calcul des taux de réduction de l'exposition	par. 157

Préparation des recommandations finales au CCCF	par. 158-164
--	--------------

Présentation des données dans les rapports des GTE au CCCF

Présentation de l'analyse des données: analyse statistique	par. 165-166
--	--------------

ANNEXES

ANNEXE I: Détermination et traitement des valeurs aberrantes/extrêmes

ANNEXE II: Nombre minimum de points de données pour l'estimation des valeurs de centile élevées

ANNEXE III: Génération de diagrammes/graphiques et de courbes se rapportant à la distribution des données d'occurrence

[ANNEXE IV : Calcul préliminaire de l'exposition alimentaire et de son taux de réduction aux contaminants d'un ou de produits de base cibles aux LM hypothétiques]

ANNEXE V: Présentation de l'analyse des données/analyse statistique

ANNEXE VI: Glossaire des termes

PRÉAMBULE

1. Les étapes de l'élaboration d'une limite maximale (LM) peuvent être les suivantes:
 - L'identification d'un nouveau problème sanitaire ou commercial relatif à une combinaison contaminant-produit de base/aliment
 - L'élaboration d'un document de travail explorant les données préliminaires sur l'occurrence, l'exposition et l'importance mondiale de la combinaison contaminant-produit/aliment.
 - L'accord du CCCF pour entreprendre de nouveaux travaux, y compris la discussion des termes de référence (TdR) d'un groupe de travail, et la soumission d'une proposition de nouveaux travaux à la CAC.
 - L'élaboration d'un document recommandant des LM dans la procédure par étapes du Codex et une analyse plus approfondie des données d'occurrence, des données d'exposition, de l'importance mondiale de la combinaison contaminant-produit de base/aliment, et de l'impact des LM proposées.
 - La recommandation d'envoyer les LM à la CAC pour adoption.
2. La principale source de données du CCCF est GEMS/Aliments, une base de données internationale gérée par l'Organisation mondiale de la santé, qui contient des données sur les taux de contaminants dans différents aliments. Les pays membres sont encouragés à soumettre des données provenant de leurs programmes nationaux de surveillance, soit de manière régulière, soit en réponse à des demandes de données de la part du CCCF; pour être soumises, les données doivent répondre à certains critères (tels que l'inclusion d'une limite de quantification (LOQ) ou d'une limite de détection (LOD) pour les données non quantifiées). Des données externes, issues par exemple de la littérature scientifique, peuvent être référencées, mais elles ne sont généralement pas utilisées pour l'élaboration de LM.
3. Avant de démarrer les travaux sur un document de travail ou un document sur les LM, le CCCF établira les TdR pour le groupe de travail et pourrait demander au secrétariat du Codex de lancer un appel de données. Comme indiqué dans le Manuel de procédure de la Commission du Codex Alimentarius (30^e édition), les TdR doivent clairement indiquer le ou les objectifs à atteindre par l'établissement du groupe de travail, la ou les langues à utiliser et le délai dans lequel les travaux sont censés être achevés. L'appel de données identifie généralement le contaminant et l'aliment/le produit de base d'intérêt, ainsi que la période pour laquelle les données sont demandées. Les appels de données précédents demandaient également des informations telles que la limite de quantification (LOQ) et la limite de détection (LOD) de la méthode d'analyse et des noms d'échantillons spécifiques; ils indiquaient également les champs de la base de données GEMS/Aliments dans lesquels les informations devaient être saisies et identifiaient la base de résultats appropriée.
4. L'établissement des TdR et la planification du champ d'application d'un appel de données constituent une étape importante du processus de collecte des données. En portant une attention particulière aux TdR et à l'appel de données, on obtiendra des données de meilleure qualité pour l'élaboration des LM.
5. La gestion des données joue un rôle clé dans le travail d'élaboration des limites maximales, et il est dans l'intérêt de tous de disposer de données de bonne qualité (telles que des données qui reflètent une image précise de la contamination des aliments, présentées à l'aide d'une analyse statistique). Les données d'occurrence devraient idéalement être obtenues par échantillonnage statistique (Réf. CXG 50-2004 Directives générales sur l'échantillonnage), et l'analyse devrait être réalisée à l'aide de méthodes validées avec des limites de quantification (LOQ) et des limites de détection (LOD) appropriées dans des laboratoires disposant de systèmes d'assurance qualité. Idéalement, les données soumises par les pays membres devraient être représentatives au niveau national.
6. L'objectif du présent document d'orientation est de fournir les éléments permettant de garantir des données de bonne qualité et d'assurer une utilisation et une analyse harmonisées des données d'occurrence disponibles par les différents GTE pour le développement/l'élaboration des LM du Codex.
7. Cette orientation est destinée à un usage interne au sein du CCCF, mais les autorités nationales/régionales peuvent utiliser les informations pertinentes contenues dans ce document d'orientation pour le développement/l'élaboration de LM nationales/régionales.

COLLECTE ET SOUMISSION DE DONNÉES

8. La page d'introduction de la base de données GEMS/Aliments de l'OMS est [Global Environment Monitoring System \(GEMS\) / Food Contamination Monitoring and Assessment Programme](#). Le processus de soumission (upload) et d'extraction (download) des données commence sur le site web de [GEMS/Food Contaminants Database](#).

9. La première page de la base de données comporte deux onglets, l'un pour la page d'accueil (*Home Page*) et l'autre pour les recherches (*Search*). Pour bénéficier de toutes les fonctionnalités, les membres doivent s'inscrire et se connecter à leur compte. Après s'être connectée, la personne désirant soumettre des données a accès à un onglet de chargement (*Upload*), qui vient s'ajouter à l'onglet Page d'accueil et à l'onglet Recherche. Elle pourra également accéder à des gabarits standard et à des gabarits pour le chargement de blocs de données, à l'outil d'apprentissage en ligne GEMS/Aliments et à des liens utiles tels que la Foire aux questions.
10. Avant de soumettre leurs données, les personnes doivent examiner les documents figurant sur le site web de GEMS/Aliments ([Nutrition and Food Safety \(who.int\)](http://Nutrition and Food Safety (who.int))) ou sur les pages GEMS/Aliments associées. Vous trouverez des instructions détaillées dans le document, INSTRUCTIONS FOR ELECTRONIC SUBMISSION OF DATA ON CHEMICALS IN FOOD AND THE DIET (ci-après dénommé «INSTRUCTIONS») sur le site web de GEMS/Aliments. Ce document fournit des instructions concernant l'enregistrement d'un compte, la connexion à la base de données GEMS/Aliments, l'insertion de données dans le gabarit Excel et le chargement du gabarit Excel. Une bonne connaissance d'Excel est très utile.
11. Des données peuvent être soumises à la base de données GEMS/Aliments à propos de n'importe quel aliment et à n'importe quel moment, et pas seulement en réponse à un appel de données spécifiant des aliments ou des périodes spécifiques présentant un intérêt particulier. Si des données sont soumises en réponse à un appel de données spécifique, pensez à noter cette information dans le champ Remarques. Les données recueillies en dehors de la période mentionnée dans l'appel de données peuvent également être soumises. Elles peuvent être instructives pour l'étude des taux de contaminants au fil du temps.
12. Si des questions se posent sur les aspects techniques de la soumission de données, la personne désirant soumettre des données doit contacter l'administrateur de la base de données de GEMS/Aliments. Les questions peuvent porter sur des messages d'erreur lors du chargement, des problèmes d'enregistrement, la manière de nommer les échantillons, les champs obligatoires, la définition des champs, les problèmes de mise en correspondance, etc.
13. Si des questions se posent à propos de la conformité des données par rapport à un appel de données spécifique, la personne désirant soumettre les données doit consulter le Président du GTE et, le cas échéant, le ou les secrétariats du Codex. Les questions peuvent porter sur le fait de savoir si les échantillons correspondent aux définitions fournies dans l'appel de données ou dans les TdR du GTE.
14. Les personnes qui soumettent des données devraient développer et conserver les métadonnées associées aux soumissions de données. Les métadonnées aideront à répondre aux questions susceptibles d'être posées par le GTE. Les métadonnées peuvent inclure l'année de production, la limite de détection (LOD) et la limite de quantification (LOQ) globales associées à un ensemble de données, des informations figurant sur les étiquettes des produits, des informations sur le lieu de collecte (par exemple, importation ou vente au détail), les noms des collaborateurs qui ont soumis les données et la date à laquelle les données ont été soumises, l'identifiant du lot associé à l'ensemble de données soumis, etc.
15. La qualité des données doit être évaluée par la personne qui les soumet avant qu'elles ne soient introduites dans GEMS/Aliments. Si des questions sérieuses se posent quant à la qualité des données (informations manquantes, analyses suspectes), ne soumettez pas les données tant que vous n'avez pas obtenu de réponses.
16. Si la personne qui soumet les données identifie un problème concernant un ensemble de données après sa soumission, consultez l'administrateur de GEMS/Aliments pour retirer ou corriger l'ensemble de données, qui devrait être identifiable grâce à l'identifiant du lot.

Compléter le gabarit GEMS/Aliments

17. La fiche du gabarit pour les soumissions standard (et non pas de blocs)¹ de données contient cinq onglets, qui comprennent (1) une liste de vérification pour les institutions qui soumettent des données, (2) le mappage alimentaire de l'échantillon, (3) un gabarit pour les résultats des analyses individuelles, (4) le système de classification de l'OMS et de FoodEx2, et (5) les produits chimiques actuellement proposés sous forme d'options dans un menu déroulant.
18. La première étape de la soumission des données consiste à remplir l'onglet «1. Début», qui contient une liste de vérification à l'intention de l'institution préparant un ensemble de données à soumettre, y compris l'identification du produit chimique d'intérêt. (Veuillez noter que pour les produits chimiques qui ne sont pas disponibles dans le menu déroulant, une option est indiquée dans les *INSTRUCTIONS*).

¹ Voir le document *INSTRUCTIONS FOR ELECTRONIC SUBMISSION OF DATA ON CHEMICALS IN FOOD AND THE DIET* pour des informations sur la soumission de gabarits pour le chargement de blocs de données.

19. La deuxième étape consiste à examiner les noms de l'aliment destiné à la consommation humaine/aliment pour les animaux dans l'ensemble de données et à associer la classification nationale des aliments avec la classification de l'OMS et de FoodEx2. L'onglet «2. Mappage de l'aliment» contient l'outil de mappage: l'identifiant local de l'aliment (colonne A, texte libre) et deux niveaux de classification dans des menus déroulants, à savoir niveau 1: Groupe OMS de l'aliment (colonne B) et niveau 2: Identifiant OMS de l'aliment (colonne C). Une fois les champs «Identifiant local de l'aliment», «Groupe OMS de l'aliment» et «Identifiant OMS de l'aliment» remplis, le code OMS de l'aliment, le code FoodEx2 et le nom FoodEx2 sont générés automatiquement dans les colonnes E, F et G de l'onglet 2.
20. Lors de la soumission de données, la fréquence à laquelle chaque aliment doit être mappé dans le gabarit de mappage de l'aliment peut être une source de confusion. Par exemple, si la personne qui soumet les données télécharge trois aliments avec les identifiants locaux d'aliments suivants: gingembre cristallisé; gingembre en poudre séché et tranches de gingembre séchées, ces trois aliments seront saisis séparément dans le gabarit de mappage de l'aliment, onglet 2, et mis en correspondance avec «Herbes, épices et condiments» (colonne B) et «Gingembre, racine» (colonne C) de l'OMS. Toutefois, si la personne qui soumet les données charge 100 points de données supplémentaires pour le «Gingembre cristallisé», la mise en correspondance ne devra être effectuée qu'une seule fois pour tous les échantillons de «Gingembre cristallisé».
21. Les *INSTRUCTIONS* précisent également que le mappage ne doit être effectué qu'une seule fois si la classification nationale est stable. Alors que certains pays ou régions peuvent avoir centralisé la soumission des données, dans d'autres pays, différentes institutions ou parties d'institutions peuvent avoir des comptes et soumettre les données séparément. Dans ce cas, les institutions doivent essayer de coordonner la manière dont elles mappent les aliments afin d'assurer la cohérence entre les différentes soumissions.
22. La troisième étape du remplissage du gabarit GEMS/Aliments consiste à saisir les résultats de l'analyse individuelle dans l'onglet «3. Résultats de l'analyse individuelle». Les champs comprennent l'Identifiant local de l'aliment (précédemment mis en correspondance avec les codes de l'onglet 2), la concentration chimique, les unités de mesure, la LOD, la LOQ, etc. Étant donné que les Identifiants locaux des aliments ont été mis en correspondance dans l'onglet 2, les colonnes B, C et D de l'onglet 3 seront remplies automatiquement avec les informations issues de la mise en correspondance. La colonne A indiquera automatiquement une erreur si l'un des champs de cet onglet est mal rempli. Les colonnes restantes doivent être remplies en suivant les instructions détaillées fournies dans le document *INSTRUCTIONS*.
23. Veuillez noter que dans le gabarit de GEMS/Aliments, les colonnes dont le titre est en bleu sont obligatoires. Les colonnes dont le titre est en blanc sont facultatives (elles peuvent être laissées vides) si l'information n'est pas disponible.
24. Les champs actuels destinés aux résultats des analyses individuelles dans la base de données GEMS/Aliments sont énumérés au Tableau explicatif 1. Le texte entre crochets (simples ou doubles) indique les changements qui ont été convenus, mais qui n'ont pas encore été mis en œuvre. Le texte bleu entre crochets simples indique les modifications basées sur les changements qui pourraient être mis en œuvre dans une première phase (le calendrier sera communiqué à la 18^e session du CCCF). Le texte violet entre doubles crochets indique les modifications qui pourraient être apportées ultérieurement.
25. Les paragraphes 26 à 52 se trouvant sous le Tableau explicatif 1 fournissent une orientation supplémentaire sur les champs à l'intention des personnes qui soumettent des données.

Tableau explicatif 1: Champs actuels du gabarit GEMS/Aliments

Colonne	Champ	Type de champ/ Menu déroulant	Obligatoire ou Facultatif	Bloc de texte (proposé nouveau ou révisé)
E	Identifiant local de l'aliment	Texte libre	Obligatoire	[Fournir un nom court, mais descriptif de l'aliment, tel que «Hoplostète orange» (par opposition à «Poisson») ou «Riz poli/blanc» (par opposition à «riz»).]
F	N° de série de l'enregistrement	Texte libre	Obligatoire	[Un numéro de série est utilisé pour chaque échantillon. Les données relatives à différents contaminants présents dans le même échantillon doivent porter le même numéro de série. Les institutions nationales devraient coordonner la sélection des numéros de série afin de s'assurer que les numéros sont informatifs et ne font pas double emploi.]
G	[Soumission] Pays/Région/Observateur	Menu déroulant (Liste des pays, régions, observateurs) - Non spécifié	[Obligatoire]	[Identifie le pays, la région ou l'observateur (région non spécifiée) qui soumet les données ; il ne s'agit pas du pays de production. Si l'observateur ne figure pas dans le menu déroulant, choisissez Non spécifié et notez le nom de l'observateur dans Remarques]

Colonne	Champ	Type de champ/ Menu déroulant	Obligatoire ou Facultatif	Bloc de texte (proposé nouveau ou révisé)
H	Contaminant	Menu déroulant (Liste des contaminants)	Facultatif	[Veuillez sélectionner un contaminant dans la liste. Un contaminant est requis, mais la saisie manuelle dans «Colonne H: Contaminant» est facultative si un contaminant a été indiqué dans la fiche 1: Début. Si «multiple » est sélectionné dans la fiche 1: Début, la saisie manuelle des contaminants dans le champ H est obligatoire.]
I	Origine de l'aliment	Menu déroulant - Production intérieure - Importation - Origine mixte - Pas d'informations	Facultatif	
J	Date d'échantillonnage	Texte libre (XXXX)	Obligatoire	
K	Représentativité de l'échantillon [fiabilité]	Menu déroulant - Échantillonnage aléatoire [(de routine)] - Échantillonnage ciblé - Pas d'informations	Obligatoire	[L'échantillonnage ciblé fait référence au suivi ciblé de découvertes spécifiques de contamination. L'échantillonnage aléatoire (de routine) fait référence à un échantillonnage non ciblé et peut inclure la surveillance de routine ou l'échantillonnage de types d'aliments spécifiques ou de pays importateurs.]
L	Identification en laboratoire	Texte libre	Facultatif	[Le laboratoire qui a effectué l'analyse]

Colonne	Champ	Type de champ/ Menu déroulant	Obligatoire ou Facultatif	Bloc de texte (proposé nouveau ou révisé)
M	Assurance qualité analytique	Menu déroulant - Assurance qualité interne et normes de référence uniquement. - Participation réussie à des tests d'aptitude/[comparaisons interlaboratoires] pertinents au cours de la période d'échantillonnage et d'analyse. - Accréditation officielle pour les méthodes pertinentes pendant la période d'échantillonnage et d'analyse. - Assurance qualité du laboratoire inconnue.	Facultatif	
N	Unités de mesure pour les taux de contaminants	Menu déroulant [• mg/kg • (µg/kg) • ng/kg • pg/kg • Bq/kg]	Obligatoire	[Contrôler soigneusement les unités. S'assurer que les unités choisies dans le menu déroulant correspondent aux résultats de l'échantillon.]
O	LOD	Texte libre	Obligatoire pour les résultats non quantifiés (c'est-à-dire non détectés)] si la LOQ n'est pas fournie. [[Facultatif]]	Saisir une valeur numérique supérieure à 0 et inférieure à la LOQ. Ce champ contient la limite de détection indiquée par le laboratoire. [La LOD et/ou la LOQ sont obligatoires si le résultat n'est pas fourni, la mention «non détecté» est inscrite dans Résultats (T).]
P	LOQ	Texte libre	Obligatoire pour les résultats non quantifiés si la LOD n'est pas fournie. [[Obligatoire]]	Saisir une valeur numérique supérieure à 0 et à la LOD. Ce champ contient la limite de quantification indiquée par [le] laboratoire. [La LOD et/ou la LOQ sont

Colonne	Champ	Type de champ/ Menu déroulant	Obligatoire ou Facultatif	Bloc de texte (proposé nouveau ou révisé)
				obligatoires si le résultat n'est pas fourni, la mention «non détecté» est inscrite dans Résultats (T).]
Q	Résultats basés sur	Menu déroulant • Teneur en graisses • Poids sec • Tel quel (cru, frais, [, tel que vendu]) • Tel que consommé	Obligatoire	
R	Portion analysée	Menu déroulant • Uniquement comestible • [Aliment entier (comestible + non comestible)]	Obligatoire	[Exemple: noix décortiquée (comestible) par rapport à noix non décortiquée (aliment entier)]
S	État de l'aliment analysé (Cuit/Cru)	Menu déroulant • Cuit • Cru • Pas d'informations	Facultatif	[Exemple: poisson cru par rapport au poisson cuit]
T	Résultats	Texte libre	Obligatoire	[La saisie d'un résultat est obligatoire: soit un zéro, soit non détecté, soit un résultat numérique. Zéro ou non détecté ne peut être indiqué que si la LOQ ou la LOD sont fournies]
U	Données individuelles par rapport aux données agrégées	Menu déroulant • Individuel • Agrégé	[[Obligatoire]]	
V	Confidentialité des données	Menu déroulant • Oui • Vide	Facultatif	[Toutes les données pour lesquelles «vide» est choisi ou aucune option n'est sélectionnée seront considérées comme non confidentielles dans le cadre du traitement et de l'analyse des données.]

Colonne	Champ	Type de champ/ Menu déroulant	Obligatoire ou Facultatif	Bloc de texte (proposé nouveau ou révisé)
W	Remarques/Références	Texte libre	Facultatif	
X	[[Année de production/récolte]]	[[Texte libre (JJJJ)]]	[[Facultatif]]	
Y	[[Informations sur la composition]]	[[Texte libre]]	[[Facultatif]]	[[Informations figurant sur les étiquettes ou déterminées de manière analytique, telles que les principaux ingrédients, la teneur en matières grasses, la teneur en eau ou le pourcentage de cacao total.]]
Z-1	[[Pays/région de production du produit fini]]	[[Menu • Pas d'informations • Pays/Régions (A-Z)]]	[[Facultatif]]	
Z-2	Pays/région d'origine des matières premières	[[Menu • Pas d'informations • Pays/Régions (A-Z)]]	[[Facultatif]]	
AA	[[Type de produit]]	[[Menu: • Destiné à une transformation ultérieure (DFP) • Prêt à consommer (RTE) • Sans objet • Pas d'informations]]	[[Facultatif]]	[[«Destiné à une transformation ultérieure» et «prêt à consommer» sont définis pour certains contaminants et produits dans la Norme générale du Codex pour les contaminants et les toxines présents dans les produits de consommation humaine et animale, CXS-193. Voir paragraphe XX ci-

Colonne	Champ	Type de champ/ Menu déroulant	Obligatoire ou Facultatif	Bloc de texte (proposé nouveau ou révisé)
				dessous.]]
BB	[[Lieu d'échantillonnage dans la chaîne de production]]	[[Menu: • Pas d'informations • Site de production • Transport de lots en vrac • Frontière (importation/exportation) • Marché/vente au détail • Autres]]	[[Obligatoire]]	[[Fournit des informations sur l'endroit où l'échantillon a été obtenu dans la chaîne de production.]]
CC	[[Principe de la méthode d'analyse]]	[[Menu • Méthode A • Méthode B • Méthode Z • Autres • Pas d'informations]]	[[Facultatif]]	

26. **E. Identifiant local de l'aliment.** Champ obligatoire. Dans la mesure du possible, la personne qui soumet les données doit fournir les noms en anglais. Ajouter des détails au nom peut aider l'analyste des données à classer l'échantillon (p. ex., «jus d'ananas-orange» par rapport à «jus»). D'un autre côté, un nom d'échantillon trop long (p. ex. la liste de tous les ingrédients d'un aliment en contenant plusieurs) peut compliquer le travail de l'analyste. Des informations supplémentaires sur le nom peuvent également être ajoutées dans la colonne Remarques.
27. **F. N° de série** Champ obligatoire. Un numéro de série (identifiant de l'échantillon) devrait être utilisé pour chaque échantillon. Si des informations sur plusieurs contaminants sont fournies pour un même échantillon, le même numéro de série doit être utilisé. Les institutions nationales sont chargées de coordonner l'utilisation du même numéro de série pour toutes les soumissions de données d'analyse concernant le même échantillon d'aliment. [Les institutions nationales devraient coordonner la sélection des numéros de série afin de s'assurer que les numéros sont informatifs et ne font pas double emploi.](#) Les conventions de numérotation doivent fournir des informations qui permettent de remonter à la personne soumettant les données, à la date de soumission ou au programme de surveillance. Éviter de numéroter avec des numéros de série peu informatifs tels que 1, 2, 3, etc. Si les soumissions sont effectuées en réponse à deux appels de données différents, la personne qui soumet le deuxième ensemble de données doit s'assurer que les nouvelles données n'ont pas remplacé les données précédentes, en vérifiant avec l'administrateur de GEMS/Aliments si nécessaire.
28. **G. [Soumission] Pays/Région/Observateur.** [\[Champ obligatoire. Identifie le pays, la région ou l'observateur \(région non spécifiée\) qui soumet les données ; il ne s'agit pas du pays de production. Si l'observateur ne figure pas dans le menu déroulant, choisissez Non spécifié et notez le nom de l'observateur dans Remarques\]](#)
29. **H. Contaminant.** [\[Champ facultatif. Un contaminant peut être ajouté sur la «feuille de travail 1»: Début» ou par saisie manuelle dans la «Colonne H: Contaminant. Un contaminant doit être identifié, mais la saisie du nom du contaminant dans la colonne H est facultative si un contaminant a été ajouté sur la fiche 1. Toutefois, si «multiple » est sélectionné dans la fiche 1: Début, la saisie manuelle des contaminants dans le champ H est obligatoire\]](#)
30. **I. Origine de l'aliment.** [\[Champ facultatif. Identifie les échantillons comme étant d'origine nationale, importée, mixte ou inconnue\]](#)
31. **J. Date d'échantillonnage.** [\[Champ obligatoire. Identifie la date d'échantillonnage.\]](#)
32. **K. Représentativité de l'échantillon.** [\[Champ obligatoire. Le terme «échantillonnage aléatoire \(de routine\)» fait référence à un échantillonnage non ciblé et peut inclure la surveillance de routine ou l'échantillonnage de types](#)

d'aliments spécifiques ou de pays importateurs spécifiques.] Par exemple, les tests effectués sur une large gamme d'échantillons importés d'une certaine catégorie d'aliments pour détecter la présence d'un contaminant donné sont «aléatoires». Le terme «échantillonnage ciblé» devrait être choisi pour les échantillonnages de suivi réalisés à la suite de constatations spécifiques de contamination. Par exemple, si un pays identifie des niveaux élevés d'un contaminant dans un échantillon provenant d'un fabricant particulier, un échantillonnage supplémentaire du même lot ou des lots produits simultanément par le même fabricant serait «ciblé».]

33. **L. [Identification du laboratoire.** Champ facultatif. Identifie le laboratoire qui a effectué l'analyse.]
34. **M. [Assurance qualité analytique.** Champ facultatif. Fournit des informations sur l'assurance qualité analytique du laboratoire d'analyse.]
35. **N. Unités de mesure des taux de contaminants.** [Champ obligatoire. Fournit des unités de mesure pour les résultats des contaminants. Contrôler soigneusement les unités. S'assurer que les unités choisies dans le menu déroulant correspondent aux résultats de l'échantillon. Veiller à ce que l'unité du rapport soit la même pour les résultats, la LOD et la LOQ. Dans le meilleur des cas, la personne qui soumet les données devrait fournir à la fois la LOQ et la LOD, même si ces champs ne sont actuellement obligatoires que pour les résultats non quantifiés..]
36. **O.LOD.** [Le champ LOD contient la limite de détection indiquée par le laboratoire. Saisir une valeur numérique supérieure à 0 et inférieure à la LOQ. Noter que la LOD ou la LOQ sont obligatoires si non détecté a été indiqué dans Résultats (T).] **[[Rapporter la LOD est facultative, mais encouragé.]]** (Commentaire: Les champs O. LOD et P. LOQ doivent être intervertis de manière à ce que le champ P. LOQ obligatoire soit placé en premier).
37. **P. LOQ.** [Le champ LOQ contient la limite de quantification indiquée par le laboratoire. Saisir une valeur numérique supérieure à 0 et à la LOD. Noter que la LOD ou la LOQ sont obligatoires si non détecté est indiqué dans Résultats (T).] **[[La LOQ est obligatoire pour les échantillons soumis après le XX/XX/XXXX. Avant le XX/XX/XXXX, la LOQ n'était obligatoire que pour les résultats non quantifiés si la LOD n'était pas fournie.]]** (Commentaire: Les champs O. LOD et P. LOQ doivent être intervertis de manière à ce que le champ P. LOQ obligatoire soit placé en premier).
38. **Q. Résultats basés sur.** Champ obligatoire. Indique si les résultats sont basés sur une analyse de la denrée «telle quelle» (crue, fraîche, telle que vendue); «telle que consommée»; ou basée sur la teneur en matières grasses ou le poids sec. Exemples: (1) «teneur en matières grasses»: résultats concernant les contaminants liposolubles dans les tissus de viande animale, basés sur la teneur en matières grasses; (2) «poids sec»: résultats concernant les contaminants dans les tissus animaux, lorsque l'échantillon est séché en laboratoire jusqu'à ce que l'humidité soit éliminée avant l'analyse; (3) «tels quels (crus, frais, tels que vendus)»: légumes et fruits secs vendus au détail; également, résultats concernant les préparations en poudre pour nourrissons, non reconstituées avec de l'eau; et (4) «telle que consommée»: résultats concernant les préparations en poudre pour nourrissons diluées.
39. **R. Portion analysée.** Champ obligatoire. Indique si l'aliment entier est analysé ou seulement la partie comestible (par exemple, noix décortiquée (comestible) ou noix non décortiquée (aliment entier)).
40. **S.État de l'aliment analysé.**[Champ facultatif. Indique si l'échantillon alimentaire est cuit ou cru (par exemple, poisson cru ou poisson cuit).]
41. **T. Résultats.** [La saisie d'un résultat est obligatoire: soit un zéro, soit non détecté, soit un résultat numérique. Zéro ou non détecté ne peut être indiqué que si la LOQ ou la LOD sont fournies.]
42. **U. Échantillon global.** Facultatif **[[Obligatoire.]]** Les données agrégées renvoient à des résultats basés sur des échantillons composites, tels que des échantillons provenant d'études de l'alimentation totale. Les données agrégées sont souvent exclues des analyses de taux d'infraction effectuées pour déterminer les limites maximales (LM) appropriées, qui sont basées sur l'observation de la distribution des données et des centiles supérieurs dépassant les LM proposées. Toutefois, des données agrégées peuvent être introduites dans la base de données GEMS/Aliments et un nombre restreint de données a été inclus dans les analyses du CCCF par le passé. L'administrateur de la base de données GEMS/Aliments ou un statisticien doit être consulté avant d'inclure des données agrégées. Si des données agrégées sont incluses dans une analyse de LM, cela doit être indiqué dans le document destiné au GTE.
43. **V. Données confidentielles.**[Champ facultatif. **Noter que toutes les données pour lesquelles «vide» a été choisi ou aucune option n'est sélectionnée seront considérées comme non confidentielles dans le traitement et l'analyse des données**] Les pays peuvent soumettre des données «confidentielles» s'ils souhaitent en limiter l'accès à la FAO, à l'OMS et aux organismes techniques apparentés, tels que le Codex. L'administrateur de GEMS/Aliments peut fournir les enregistrements portant la mention «confidentiel» aux présidents des GTE; par conséquent, avant de télécharger des données, les présidents des GTE doivent toujours consulter l'administrateur de GEMS/Aliments à propos de l'extraction des données. Si un pays a soumis des données

portant la mention «confidentiel» en réponse à un appel de données, il doit également en informer le Président du GTE lors de la phase d'extraction/d'analyse des données.

44. **W. Remarques/Références.** [Champ facultatif qui peut être utilisé pour ajouter des informations pertinentes pour lesquelles il n'existe pas de champ défini. Les informations figurant sur les étiquettes des produits (telles que les ingrédients ou les noms détaillés des produits) ou les informations détaillées sur la méthode d'analyse sont des exemples d'informations qui ont été incluses dans cette colonne. Une référence à un appel de données spécifique est une autre information qui peut être inscrite dans cette colonne.]
45. **[[X. Année de production/récolte.** Champ facultatif. Fournit des informations sur l'année de production, par exemple pour les denrées alimentaires d'origine animale, ou sur l'année de récolte, par exemple pour les denrées alimentaires d'origine végétale, si l'information est disponible.]]
46. **[[Y. Informations sur la composition.** Facultatif. Informations figurant sur **les étiquettes**, telles que les principaux ingrédients, la teneur en matières grasses, la teneur en eau ou le pourcentage de cacao total pour le chocolat.]]
47. **[[Z. Pays/région d'origine/de production.** Facultatif. Fournit des informations sur le pays ou la région d'origine/de production de [l'aliment ou de la matière première ou du lot ou de l'envoi échantillonné]. Si cette information n'est pas connue, il est possible de sélectionner Inconnu.]]
48. **[[AA. Type de produit** Facultatif. Indique si l'échantillon est «destiné à une transformation ultérieure» ou «prêt à consommer», ou cette information est « Sans objet» ou «Inconnue». «Destiné à une transformation ultérieure» ou «prêt à consommer» sont définis à plusieurs endroits dans la [Norme générale du Codex pour les contaminants et les toxines présents dans les produits de consommation humaine et animale](#) (CXS-193-1995). Pour l'aflatoxine dans le maïs, le sorgho, les fruits à coque, les arachides et les figues sèches ou le déoxynivalénol dans les céréales, «destinés à une transformation ultérieure» signifie destinés à subir une transformation/un traitement supplémentaire qui s'est avéré réduire les niveaux d'aflatoxine ou de déoxynivalénol avant d'être utilisés comme ingrédients dans les denrées alimentaires, transformés d'une autre manière ou proposés à la consommation humaine. Pour les aflatoxines dans les fruits à coque et figues sèches, «prêt à consommer» signifie qu'ils ne sont pas destinés à subir une transformation/un traitement supplémentaire qui permettrait de réduire les concentrations d'aflatoxines avant d'être utilisés comme ingrédient dans les produits alimentaires, autrement transformés ou proposés pour la consommation humaine. Pour les produits de base mentionnés ci-dessus, il est possible de sélectionner «destiné à une transformation ultérieure» et «prêt à consommer» ou «inconnu»; pour les autres produits, il convient de sélectionner «sans objet».
49. La CXS-193-1995 fait également référence à plusieurs endroits à des repas pour nourrissons «prêts à consommer» et à du lait liquide «destiné à une transformation ultérieure». Le champ AA ne s'applique pas à ces utilisations de «prêt à consommer» et «destiné à une transformation ultérieure».]
50. **[[BB. Lieu d'échantillonnage dans la chaîne de production.** Obligatoire. Fournit des informations sur l'endroit où l'échantillon a été obtenu dans la chaîne de production, à savoir: Site de production (par exemple, entreprise, usine alimentaire), transport de lots en vrac, frontière (par exemple, test à l'importation ou à l'exportation), marché/vente au détail, ou autre. Si cette information n'est pas connue, il est possible de sélectionner Inconnu. Si l'option Autre est sélectionnée, des informations peuvent être ajoutées aux Remarques.]]
51. **[[CC. Principe de la méthode d'analyse.** Facultatif. La méthode d'analyse, telle que par ICPMS ou par GC-MS, peut être sélectionnée à partir d'un menu déroulant.]]
52. **Erreurs.** Avant d'introduire des données, la personne qui les soumet doit relire le fichier attentivement afin de déceler d'éventuelles erreurs. Pendant le chargement, le fichier est vérifié pour identifier les problèmes avant de saisir des données dans la base de données. Il incombe à la personne qui soumet les données de corriger les erreurs, puis de soumettre à nouveau le gabarit. Des ensembles de données peuvent être rejetés pour diverses raisons, dont certaines sont énumérées ci-dessous. L'administrateur de la base de données GEMS/Aliments peut être contacté pour obtenir de l'aide.
 - a. Résultat rapporté < LOD, - LOQ ou LOD manquante lorsque le résultat est **non détecté** **[[LOQ manquante-]], LOD > LOQ rapporté**
 - b. Dates introduites dans un format incorrect
 - c. Champs obligatoires incomplets
 - d. Doublons dans la fiche en cours de traitement ou dans la base de données

EXTRACTION DE DONNÉES

53. Le processus d'extraction des données commence sur le site web de la base de données: [GEMS/Food](#)

[Contaminants Database](#). Comme indiqué plus haut, pour bénéficier de toutes les fonctionnalités, les membres doivent s'inscrire et se connecter à leur compte. Après s'être connectés, les analystes verront s'ouvrir une page comportant deux onglets, l'un pour la page d'accueil (*Home Page*) et l'autre pour les recherches (*Search*). L'onglet Page d'accueil contient un petit nombre d'ensembles de données extraits par région et par contaminant. Pour effectuer des recherches spécifiques, l'analyste sélectionne l'onglet Recherche. La fonction de recherche permet à l'analyste de filtrer les données par région de l'OMS, par contaminant, catégorie d'aliment, nom de l'aliment et période d'échantillonnage. Ces filtres lui permettent d'identifier les données répondant à un appel de données ou à des TdR particuliers.

54. Afin d'identifier l'ensemble de données le plus précis à extraire pour le développement des propositions de LM, il est préférable de consulter l'administrateur de la base de données GEMS/Aliments. Lors de la soumission des données, les utilisateurs peuvent faire des choix qui risquent d'entraîner l'omission de données lors de l'extraction. Un exemple: des données téléchargées en tant qu'«aliments pour nourrissons et enfants» peuvent ne pas être prises en compte lors d'une recherche limitée à des «jus de fruits et de légumes». Un autre exemple: des données sur les jus peuvent être associées par erreur aux «fruits et produits à base de fruits», alors que l'identifiant local de l'aliment ou le champ Remarques identifie clairement les échantillons comme étant des jus. La consultation de l'administrateur de base de données de GEMS/Aliments avant l'extraction des données peut aider le GTE à s'assurer qu'il a extrait de GEMS/Aliments toutes les données pertinentes pour l'analyse des LM.
55. Les données confidentielles sont une autre raison pour laquelle, avant de télécharger des données, les présidents des GTE doivent toujours consulter l'administrateur de GEMS/Aliments. L'administrateur de GEMS/Aliments peut fournir les enregistrements portant la mention «confidentiel» aux Présidents des GTE. Ces enregistrements n'apparaîtront pas lors d'une recherche de routine telle que décrite ci-dessus. Les membres des GTE intéressés par une analyse plus détaillée des données confidentielles peuvent consulter le président du GTE.
56. Il est important de conserver un enregistrement de tous les filtres et termes de recherche pour le rapport du GTE.
57. *[[Des modifications ont été apportées à la base de données GEMS/Aliments en 2025/6 afin de rendre la LOQ obligatoire et d'introduire de nouveaux champs tels que X, Y, Z1, Z2, AA, BB, CC. Les données transmises avant ces changements doivent être considérées comme valides et être utilisées pour le développement des LM.]]*

SELECTION/NETTOYAGE DE DONNEES

CONSIDÉRATIONS GÉNÉRALES

58. Cette section fournit des conseils sur la sélection et le nettoyage des données à soumettre et à analyser dans le cadre des travaux du CCCF. Elle vise à favoriser un traitement cohérent et transparent des données provenant de différentes sources, tout en reconnaissant que les décisions clés — par exemple, si les LM doivent être élaborées — relèvent de la compétence du CCCF. Les procédures décrites ici visent à faciliter l'analyse de données de haute qualité tout en promouvant l'harmonisation entre les contributeurs.

Origine des données

59. Pour l'élaboration des LM basées sur les données d'occurrence, seules les données de la base de données GEMS/Aliments seront utilisées. Les données non GEMS/Aliments ne peuvent être utilisées qu'à l'appui d'une analyse complémentaire visant à mieux comprendre les données, par exemple lorsque la base de données GEMS/Aliments ne contient que des données limitées pour certaines périodes ou régions, en particulier des données limitées provenant de pays producteurs primaires.
60. Les données non GEMS/Aliments, telles que les données directement soumises au GTE par le(s) pays/organisation membre(s) du Codex ou le(s) observateur(s), ou obtenues par le biais d'une recherche documentaire, sont également soumises à des procédures de nettoyage, si nécessaire.

Nettoyage des données

61. L'objectif du nettoyage des données est de supprimer les échantillons non pertinents ou inappropriés de l'ensemble de données avant d'analyser l'ensemble de données pour recommander des LM.

62. Exemples d'échantillons devant être supprimés d'un ensemble de données:
- Échantillons qui sont clairement en dehors des TdR des travaux, par exemple des échantillons de ketchup pour des travaux sur la sauce tomate.
 - Échantillons dont la date ne correspond pas à celle des TdR des travaux, par exemple, des échantillons datant de 20 ans et dont les TdR font référence à des données datant de 10 ans. Ceci est particulièrement important si un code d'usages pour la prévention et la réduction de la contamination des contaminants pertinents dans les aliments a été adopté depuis que les anciennes données ont été générées.
 - Échantillons manquant d'informations cruciales (voir paragraphes 66-73 ci-dessous).
 - Échantillons dont la LOQ est inacceptablement élevée (voir paragraphes 93-96 ci-dessous).
63. Le nettoyage des données ne concerne que l'ensemble des données extraites. Les données originales de la base de données GEMS/Aliments ne seront pas modifiées et ne seront pas affectées par les étapes indiquées ci-dessous, à moins que la personne soumettant les données ne demande des corrections ou d'autres modifications à l'administrateur de GEMS/Aliments.
64. Pour le nettoyage des données, il est recommandé de faire appel à un expert du contaminant en question, qui peut avoir une idée de la nature irrégulière ou non des tendances observées dans les données.
65. Toutes les étapes du nettoyage des données devraient être enregistrées et décrites dans le document de travail final du Codex présenté par le GTE à la session plénière du CCCF, par exemple, des informations détaillées sur la raison des exclusions de données (par exemple, des LOD/LOQ spécifiées plus élevées que les LM hypothétiques envisagées, des valeurs aberrantes douteuses, etc.), le nombre de points de données exclus durant le processus de nettoyage, si possible aussi en incluant une ventilation du nombre d'exclusions à chaque étape, etc.

Traitement des données

Informations manquantes

66. Une fois que toutes les données non pertinentes ou inappropriées ont été supprimées de l'ensemble de données, les données ne doivent pas être exclues si tous les champs obligatoires sont remplis (voir la section sur la collecte et la soumission des données) et si les données répondent aux critères de téléchargement dans la base de données GEMS/Aliments. Il convient de noter que même si tous les champs obligatoires sont remplis et que les données répondent aux critères, les données peuvent encore être incomplètes pour l'obtention des LM.
67. Dans les cas où l'information sur l'échantillon dans la base de données GEMS/Aliments est nécessaire, mais incomplète (par exemple, manquante ou peu claire), la première étape devrait être de contacter le point de contact du pays/organisation ou de l'observateur qui soumet les données pour permettre d'obtenir l'information manquante. Ceci est particulièrement important si les informations manquantes ont été demandées dans le cadre d'un appel de données. L'administrateur de la base de données GEMS/Aliments peut également être chargé de cette tâche.
68. Si des informations manquantes sont disponibles, un fichier de données corrigé doit être fourni par le pays/organisation ou l'observateur au GTE et à l'administrateur du GEMS/Aliments. L'analyse à l'aide des échantillons corrigés pourra se poursuivre. La demande et les corrections doivent être notées dans le document de travail du GTE.
69. Si les informations manquantes ne peuvent être obtenues et que le président du GTE conclut que les données doivent être exclues de l'analyse en raison d'informations manquantes, le GTE doit noter l'exclusion des données et les impacts possibles de l'exclusion dans son document de travail.
70. Exemples d'informations manquantes indiquant que les données devraient éventuellement être exclues d'une analyse plus approfondie:
- l'état de l'aliment analysé n'est pas précisé (par exemple, cuit par rapport à cru, alors que l'analyse est censée porter uniquement sur des aliments crus)
 - description adéquate du produit dans le champ de l'Identifiant local de l'aliment (par exemple, l'analyse porte sur le «maquereau», mais le produit est décrit comme étant du «poisson», et l'analyse dépend de l'identification de l'espèce de poisson)
 - et autres
71. Exemples d'informations manquantes qui n'empêcheraient pas la poursuite de l'analyse des données (en fonction de l'examen au cas par cas):
- informations sur l'échantillonnage: type d'échantillonnage, lieu d'échantillonnage dans la chaîne de

- production
- l'état du produit, par exemple cru ou cuit, n'est pas indiqué dans le champ «S», mais l'information peut être déduite d'autres informations fournies, par exemple l'échantillon est décrit comme du poisson cuit.
- principe de la méthode d'analyse utilisée
- lorsque la LM est basée sur une somme de composants et que les données ne sont pas déclarées pour tous les composants, mais pour ceux qui contribuent de manière significative à la somme ou lorsque l'occurrence est rapportée comme une somme.

72. Le président du GTE ne doit pas exclure toutes les données pour lesquelles il manque un champ facultatif, mais doit examiner les détails nécessaires au développement/à l'élaboration des LM. Par exemple, pour certains produits de base, des informations supplémentaires peuvent être nécessaires pour élaborer les LM. Ainsi, pour les céréales brutes, des informations sur l'étape de transformation peuvent être importantes pour proposer des LM, mais ces informations ne seraient pas nécessaires pour les boissons produites en tant que produits finis.
73. Comme indiqué précédemment, les données non GEMS prises en compte dans le cadre d'une analyse complémentaire peuvent manquer d'informations cruciales. Le président du GTE doit appliquer les mêmes critères et questions (par exemple, l'information manquante est-elle essentielle à l'inclusion) aux données non GEMS.

Informations erronées

74. Dans les cas où il y a des indications claires que l'unité des données ou la base sur laquelle les données sont rapportées sont erronées, la première étape devrait être de contacter le point de contact pour le pays/l'organisation ou l'observateur qui a soumis les données et de leur demander de revoir les entrées pour procéder à d'éventuelles corrections. L'administrateur de la base de données GEMS/Aliments peut également être chargé de cette tâche.
75. Si le pays/organisation ou l'observateur qui soumet les données convient qu'une correction est nécessaire, le pays ou l'observateur qui soumet les données doit fournir un fichier de données corrigé au GTE et à l'administrateur de la base GEMS/Aliments. L'analyse à l'aide des échantillons pourra se poursuivre. La demande et les corrections doivent être notées dans le document de travail du GTE.
76. Si l'exactitude des données ne peut être confirmée et que des corrections ne peuvent être apportées, ces données doivent être exclues de toute analyse ultérieure.
77. Exemples d'erreurs apparentes qui devraient conduire à contacter les personnes soumettant les données pour qu'elles les corrigent et les soumettent à nouveau:
 - Toutes les données d'un ensemble de 200 échantillons sont exprimées en µg/kg, à l'exception de 5 points de données quantifiés exprimés en mg/kg. Lorsque l'on trace ces données dans une courbe de distribution de fréquences, après les avoir converties dans la même unité, les 5 points de données en mg/kg sont identifiés comme d'éventuelles valeurs aberrantes (voir paragraphes 82-92 et l'Annexe I).
 - 195 résultats dans un ensemble de données de 200 échantillons de denrées alimentaires ayant une teneur en matières grasses typique de 5 % se situent dans la fourchette de 0 à 20 mg/kg; toutefois, 5 points de données se situent dans la fourchette de 100 à 400 mg/kg, ce qui suggère qu'ils ont été rapportés sur la base des matières grasses plutôt que sur la base du poids entier désigné. Lorsque l'on trace ces données dans une courbe de distribution de fréquences, elles sont identifiées comme d'éventuelles valeurs aberrantes (voir paragraphes 82-92 et Annexe I).

Données pour lesquelles l'information sur la partie analysée n'est pas claire

78. Pour certains aliments (par exemple, les fruits, le riz), si la portion analysée n'est pas claire (par exemple, fruit pelé vs fruit entier, ou riz décortiqué vs riz poli), le point de contact du pays/de l'organisation ou de l'observateur qui a soumis les données peut être contacté pour obtenir des éclaircissements. Si aucun éclaircissement n'est fourni, il convient de déterminer si les informations peu claires sont importantes ou pertinentes pour le contaminant en question et la concentration finale trouvée dans le produit. En outre, pour certaines denrées alimentaires, on peut supposer que la portion a été analysée dans l'état où elle est habituellement vendue/consommée, par exemple les agrumes sont généralement frais, sauf s'ils sont clairement identifiés comme étant en conserve. Toute hypothèse de ce type doit être enregistrée et présentée dans le document final du GTE à la session plénière du CCCF. Si aucune hypothèse raisonnable ne peut être formulée, ces données doivent être exclues de toute analyse ultérieure, à moins que les informations nécessaires n'aient été obtenues.

Données provenant d'échantillons soupçonnés d'être frauduleux/économiquement falsifiés

79. Pour déterminer si les échantillons contaminés sont le résultat d'une falsification frauduleuse/économique, il

faut d'abord tenir compte de la nature du contaminant (par exemple, plomb ou mycotoxine). Des différences importantes d'une année à l'autre peuvent être le résultat de la variabilité naturelle (par exemple, un niveau élevé de mycotoxines dû à des conditions climatiques spécifiques dans une certaine région/année de production). Dans d'autres cas, les écarts importants peuvent être le résultat de mauvaises pratiques (par exemple, le plomb). Les signes possibles d'échantillons frauduleux/économiquement falsifiés sont les suivants:

- certains échantillons sont des ordres de grandeur plus élevés que d'autres, par exemple 0,1 mg/kg par rapport à 100 mg/kg, ou
- la variabilité temporelle des données (selon le contaminant), par exemple, les données sont beaucoup plus élevées pour une année de l'ensemble de données.

Les données qui sont clairement liées à des échantillons frauduleux ou économiquement falsifiés doivent être exclues de l'analyse et l'exclusion doit être documentée.

Données provenant d'un échantillonnage ciblé et biais

80. L'échantillonnage ciblé diffère de l'échantillonnage aléatoire en ce sens qu'il s'agit d'un suivi ciblé de découvertes spécifiques de contamination. En principe, ces données ne devraient pas être utilisées pour calculer les LM, car elles ne sont pas représentatives de la population générale et peuvent ne pas refléter les niveaux réalisables dans des situations normales.
81. Il convient également de noter qu'un certain biais pourrait apparaître lors d'un échantillonnage aléatoire, car il peut y avoir des raisons de procéder à un échantillonnage plus poussé dans des régions ou des types de produits spécifiques. Ces données peuvent inclure des niveaux supérieurs ou inférieurs à la fourchette normale et ne doivent pas être exclues sans autre considération, car elles reflètent la variation naturelle des données d'occurrence.

Détermination et traitement des valeurs aberrantes/extrêmes

82. Le président du GTE doit examiner l'ensemble des données pour déterminer s'il existe des valeurs aberrantes/extrêmes qui devraient être supprimées. Cette section fournit des orientations sur les différentes approches permettant d'identifier les valeurs aberrantes/extrêmes et les protocoles pour les supprimer, le cas échéant.
83. En règle générale, il est important que les valeurs aberrantes/extrêmes ne soient pas écartées, à moins qu'il n'y ait une raison valable de le faire.
84. Les valeurs extrêmes peuvent être valables, par exemple en raison de l'hétérogénéité de la distribution des contaminants (comme les points chauds pour les mycotoxines) ou en raison de la variation naturelle des contaminants mesurés (par exemple, en raison des conditions météorologiques, de l'état du sol, etc.) D'autres valeurs extrêmes peuvent être erronées, par exemple des erreurs de mesure et de traitement des données, y compris des calculs incorrects ou l'utilisation d'une unité de mesure incorrecte, ou résulter d'un comportement frauduleux (falsification économique).
85. Les valeurs erronées ou les valeurs résultant d'une falsification économique doivent toujours être supprimées de l'ensemble de données final avant de déterminer les LM. Les valeurs extrêmes valides devront être examinées au cas par cas, en utilisant d'abord l'inspection visuelle des données (voir Annexe III pour des exemples), puis les méthodes statistiques (Annexe I).
86. La présence de valeurs aberrantes dans les ensembles de données a un impact significatif sur la moyenne arithmétique et les valeurs extrêmes, mais pas sur la médiane. Toutefois, il convient de tenir compte du pourcentage de valeurs aberrantes potentielles présentes dans l'ensemble des données. Étant donné que les valeurs du centile élevées, plutôt que les valeurs maximales, servent de base à l'élaboration des LM fondées sur les taux de rejet, l'impact des valeurs aberrantes sur les LM dérivées sera généralement faible. Cependant, dans les cas où un pourcentage notable de points de données (par exemple 2-5 %) sont exclus de l'ensemble de données en tant que valeurs aberrantes, cela pourrait affecter les valeurs de centile élevées (voir également le paragraphe 92). Dans ces cas, il convient de fournir un commentaire concernant l'effet de l'exclusion des valeurs aberrantes sur la possibilité d'atteindre les LM (c'est-à-dire le taux de rejet) considérées.
87. L'Annexe I du présent document d'orientation fournit plus de détails sur les approches statistiques permettant d'identifier les valeurs aberrantes.
88. Le terme «valeur aberrante» défini dans la Directive sur la Terminologie analytique (CXG 72-2009) suppose une distribution normale dans un ensemble de données composé de résultats d'analyses répétées du même échantillon. Elle stipule que *«les valeurs statistiques aberrantes sont éliminées à moins que le statisticien ne décide, pour de bonnes raisons, de les conserver»*. En revanche, les ensembles de données traités dans ce document d'orientation sont des résultats d'analyse provenant d'une variété d'échantillons et de différentes méthodes d'analyse. La distribution de ces données étant inconnue (dans de nombreux cas, elle ne suit pas une

distribution normale) et les données pouvant être combinées à partir de sources multiples, il est difficile de prévoir l'étendue de la variation au sein d'un ensemble de données. Par conséquent, ce document d'Orientation recommande que les valeurs statistiques aberrantes ne soient pas écartées à moins qu'une bonne raison de les exclure ne soit identifiée et expliquée scientifiquement.

89. Les valeurs extrêmes identifiées comme des erreurs doivent être traitées comme indiqué aux paragraphes 74 à 77. Toutefois, les valeurs extrêmes qui n'ont pas d'explication claire doivent être conservées et évaluées afin de déterminer si elles doivent être traitées comme des valeurs aberrantes dans l'ensemble de données final. Une analyse de sensibilité peut être réalisée pour comprendre l'impact de l'inclusion des valeurs aberrantes sur l'évaluation globale.
90. Étant donné que les valeurs extrêmes peuvent avoir de nombreuses causes et que certaines de ces valeurs peuvent ne pas être considérées comme extrêmes si elles sont combinées avec des données provenant d'autres sources (pays/régions, années différentes, etc.), l'exclusion éventuelle d'une valeur extrême en tant que valeur aberrante doit être déterminée sur la base de l'ensemble de données combinées et nettoyées. Si des ensembles de données individuels sont analysés séparément, il convient d'accorder une plus grande attention à l'exclusion des valeurs extrêmes en tant que valeurs aberrantes.
91. Dans certains cas, les valeurs extrêmes sont scientifiquement valables en fonction des conditions de production et météorologiques, et d'autres facteurs potentiels tels que les éruptions volcaniques, etc. Compte tenu des caractéristiques de la distribution des contaminants des données d'occurrence dans les aliments, il n'est pas recommandé d'exclure simplement les valeurs extrêmes sur la base des résultats des tests statistiques de valeurs aberrantes ou d'autres méthodes telles que l'inspection visuelle. Étant donné que la gamme des concentrations et des schémas de distribution qui peuvent être empiriquement ou théoriquement supposés varie considérablement en fonction du type de contaminant (métaux lourds, mycotoxines, etc.), le traitement des valeurs extrêmes doit être déterminé au cas par cas. Par exemple, une attention particulière doit être accordée aux mycotoxines dont les concentrations peuvent varier de manière significative en fonction des méthodes d'échantillonnage utilisées, en raison de la distribution hétérogène dans un lot, ainsi que des variations annuelles très importantes.
92. Si les valeurs aberrantes sont exclues, il est recommandé que la raison de l'exclusion soit clairement indiquée dans le document final présenté par le GTE à la session plénière du CCCF. Des analyses de sensibilité peuvent être effectuées pour montrer comment l'exclusion ou la non-exclusion des valeurs aberrantes peut ou non affecter le calcul des valeurs de centile élevées. Il convient de rappeler que quelques valeurs extrêmes restant dans l'ensemble de données n'auront que peu d'effet sur le calcul des valeurs de centile élevées, à condition que le nombre total de points de données dans l'ensemble de données soit suffisamment supérieur au nombre minimum de points de données requis pour calculer les valeurs de centile élevées.

Limite de quantification (LOQ) et Limite de détection (LOD)

Exclusion sur la base de la LOQ

93. Différentes méthodes d'analyse fournissent des LOD et des LOQ différentes. Une LOQ élevée ne signifie pas automatiquement que les données doivent être exclues.
94. Orientation pour l'inclusion/exclusion de données dans différents scénarios de LOQ/LOD
 - Dans le cas où aucune LOQ/LOD n'est fournie pour un ensemble de données spécifique
 - Le pays/organisation ou l'observateur qui soumet les données doit être contacté dans un premier temps pour obtenir ces informations (c.-à-d. LOD et/ou LOQ).
 - Dans le cas où l'ensemble de données contient (presque) tous les résultats quantifiés: l'ensemble de données peut être utilisé.
 - Lorsque l'ensemble de données contient une quantité significative de données censurées à gauche (c'est-à-dire des données individuelles sans valeurs quantifiées ou finies, généralement appelées données inférieures aux LOQ/LOD rapportées): l'ensemble de données ne doit pas être utilisé.
 - Dans le cas où une LOQ est fournie:
 - Identifier un seuil pour la LOQ dans l'analyse en fonction des LM considérées (exemples: $LOQ < LM$ en cours de discussion, $LOQ < 1/2 LM$ en cours de discussion).
 - Dans le cas où l'ensemble de données contient une quantité importante de données censurées à gauche: des informations contextuelles supplémentaires doivent être prises en compte (par exemple, les données proviennent-elles d'un pays importateur ou d'un pays producteur).

95. Si la quasi-totalité des données de l'ensemble de données sont inférieures à la LOQ ou déclarées comme non détectées (ND, <LOD), il n'est pas possible d'estimer des valeurs de centile élevées pour établir les LM proposées. Lorsqu'il n'y a qu'un petit nombre de valeurs quantitatives, l'ensemble de données doit être traité au cas par cas en suivant les conseils fournis dans la section «Traitement des ensembles de données comportant une grande proportion de données censurées à gauche». Dans ce cas, lorsque l'on propose des LM, il n'est pas approprié de calculer des valeurs de centile élevées en utilisant uniquement des valeurs quantitatives, car cela peut aboutir à des LM inutilement élevées.
96. Le document de travail du GTE devrait indiquer clairement les critères selon lesquels certaines données ont été exclues de l'ensemble de données en raison de LOD/LOQ élevées (par exemple, la LOD/LOQ rapportée pour certains échantillons est supérieure à la LM proposée, ou la LOD/LOQ rapportée pour certains échantillons est supérieure de «x» ordres de grandeur à la LOD/LOQ la plus basse de la grande majorité des autres échantillons de l'ensemble de données, ce qui suggère une erreur) ou si l'ensemble de données doit être exclu de l'analyse, car l'élimination de données individuelles peut introduire un biais.

Somme des composants et LOQ

97. Dans le cas de niveaux de contaminants qui sont une somme de composants (par exemple, les aflatoxines totales), les éléments suivants doivent être pris en compte.
- La règle générale est que les niveaux de contaminants qui sont une somme de composants sont rapportés comme la limite inférieure, c'est-à-dire que pour les niveaux non quantifiés de composants, les valeurs sont fixées à 0.
 - Les LOD ou les LOQ des composants individuels doivent être fournies pour les composants non quantifiés en dessous de la LOD ou de la LOQ.
 - Lorsque seules des données sur des composants individuels sont communiquées, les données individuelles peuvent être additionnées pour obtenir un résultat total.
 - Dans certains cas, il peut être approprié de déclarer les taux de contaminants qui sont une somme de composants en utilisant une approche de limite moyenne (c'est-à-dire que les niveaux non quantifiés des composants sont fixés à une ½ LOQ (ou à la LOD si la LOQ n'est pas fournie) ou une approche de limite supérieure (c'est-à-dire que les résultats non quantifiés sont fixés à 1 LOQ (ou à la LOD si la LOQ n'est pas fournie); toutefois, ces cas doivent être clairement identifiés à l'avance avant la soumission des données dans les instructions relatives à l'appel de données.

ANALYSE DES DONNÉES: GÉNÉRATION D'UN APERÇU DE DONNÉES

Aperçu des pays, nombre de points de données, période de couverture des données

98. Après le nettoyage de l'ensemble des données, les données restantes sont considérées comme étant d'une qualité suffisante pour l'analyse. Un aperçu de ces données restantes avec des détails (par exemple, le pays de production du produit fini, l'année de production/récolte, la quantité de données incluses et exclues) doit être fournie dans un tableau et décrite dans le document de travail. Toutes les mesures prises dans le cadre du nettoyage des données, et la justification et les hypothèses formulées doivent être fournies avec l'aperçu. En outre, il pourrait être utile de fournir des informations (par exemple, de la FAO) sur les principales régions de production du produit de base en question. Sur la base de cette vue d'ensemble, le GTE peut également présenter une analyse plus ciblée de zones géographiques et de périodes spécifiques.

Couverture géographique des données d'occurrence fournies

99. Lors de l'évaluation d'une LM pour un produit de base particulier, l'ensemble de données doit inclure une représentation des régions de production qui sont importantes pour le commerce international. Par conséquent, il est utile (et peut être exigé à l'avenir) que le pays/la région qui produit les produits finis soit indiqué dans la base de données GEMS/Aliments (voir la section «Collecte et transmission des données»). De même, les principales régions de production pour certains produits peuvent être limitées géographiquement; si c'est le cas, le président du GTE peut rapporter cette information dans le document de travail. Dans ce contexte, les données des régions productrices doivent être examinées par rapport aux données des pays qui importent les denrées alimentaires, car ces données peuvent être faussées si les aliments doivent se conformer aux exigences du pays importateur, comme une LM déjà établie dans ce pays. Si une région a des LM très restrictives, de sorte que les taux de contamination des aliments importés sont biaisés vers la gauche, il se peut que cela ne soit pas une représentation exacte de la variabilité de la contamination des régions productrices. Toutefois, les données des pays importateurs reflètent également les aliments (ingrédients) tels qu'ils sont échangés au niveau international et tels qu'ils sont consommés, et devraient être prises en compte dans une certaine mesure. En effet, une

contamination supplémentaire pourrait avoir lieu pendant le transport depuis le pays producteur (par exemple, production de mycotoxines).

100. Dans certains cas, il pourrait être approprié de donner la priorité aux ensembles de données des pays producteurs par rapport aux ensembles de données des pays importateurs. Toutefois, dans ce cas, il convient de garantir que les ensembles de données provenant des pays producteurs reflètent bien la mise en œuvre des bonnes pratiques agricoles et de fabrication prévues par les Codes d'usages du Codex et sont représentatifs des produits qui feraient l'objet d'un commerce international.
101. Une autre option consiste à fournir des analyses distinctes pour la production et l'importation d'ensembles de données par pays. Si possible, une analyse de sensibilité sur l'utilisation des données des pays producteurs par rapport aux pays importateurs devrait être réalisée pour guider la sélection des données pour l'élaboration de LM.
102. Ce n'est que s'il y a suffisamment de données indiquant des différences importantes dans les taux déclarés entre les régions ou entre les pays d'une région que l'analyse peut être effectuée par région ou par pays. Il convient de noter, dans le cas d'une approche par pays, que cette démarche doit être effectuée pour les principaux pays producteurs de la région et que des données suffisantes doivent être disponibles. (Voir la section sur l'Analyse statistique des données d'occurrence/Traitement des ensembles de données pour l'élaboration des LM).
103. En résumé, bien qu'il existe différentes approches possibles pour examiner les données des pays importateurs par rapport aux pays producteurs ou aux données régionales, il est important de garder à l'esprit que les LM du Codex sont des normes mondiales, et que l'approche par défaut pour le traitement des données devrait donc être d'analyser les données au niveau mondial.
104. Orientation pour les ensembles de données qui manquent de couverture géographique:
 - Si la/les région(s) pour laquelle/lesquelles les données manquent est/sont une/des région(s) de production importante(s) et à condition qu'elle(s) s'engage(nt) clairement à fournir des données supplémentaires, quelques années supplémentaires (par exemple 2 à 3 ans) sont généralement accordées pour la collecte de données avant de poursuivre la discussion sur les propositions de LM, à condition qu'il n'y ait pas d'urgence et que le CCCF en convienne. Après l'expiration des années supplémentaires accordées, la discussion sur les LM pourrait être poursuivie sur la base des données disponibles, que la couverture géographique ait été atteinte ou non.
 - Si la ou les régions productrices importantes ne s'engagent pas à fournir les données supplémentaires ou si la collecte de données supplémentaires dans les délais convenus (par exemple, 2 à 3 ans) n'est pas possible, l'examen des LM se poursuivra sur la base des données disponibles ou sera interrompu.
 - Si la/les région(s) pour laquelle/lesquelles les données sont insuffisantes n'est/sont pas une/des région(s) de production importante(s): l'examen sur les LM se poursuivra sur la base des données disponibles.
 - S'il est urgent d'établir une LM pour la protection de la santé des consommateurs, un compromis doit être trouvé au sein de la CCCF pour fixer une LM sur la base des données disponibles. Dans ce cas, la LM peut être révisée dans un délai de 3 à 5 ans afin de déterminer si des ajustements sont nécessaires lorsque davantage de données seront disponibles.

Période couverte par les données d'occurrence fournies

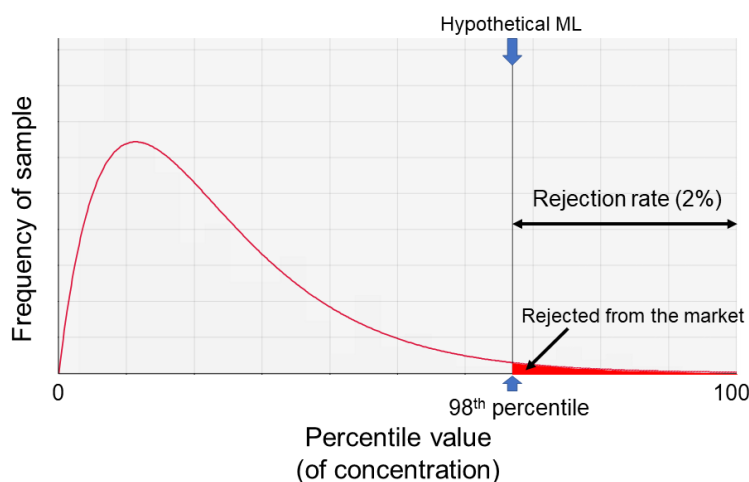
105. Il convient que les données d'occurrence fournies se rapportent à plusieurs années de production pour l'élaboration de la LM. Les exigences peuvent être différentes selon les types de contaminants (par exemple, mycotoxines, toxines végétales, biotoxines marines, contaminants liés à la transformation et à l'environnement) et sont fonction de la variation annuelle supposée ou de l'évolution de la contamination dans le temps.
106. Pour les contaminants tels que les mycotoxines, dont on sait qu'ils varient d'une année à l'autre, les données des dix dernières années peuvent fournir une très bonne représentation de la variation d'une année à l'autre; toutefois, dans certains cas, il peut être nécessaire de prendre en compte plus de 10 années de données (par exemple, si l'effort d'échantillonnage a été réduit au cours des dernières années ou si l'on dispose de moins d'ensembles de données de meilleure qualité). Pour d'autres contaminants, la variation d'une année sur l'autre est moins importante et des données plus récentes (ou une période plus courte) peuvent être sélectionnées. En tout état de cause, il convient de déterminer si les données datant de plus de 10 ans sont pertinentes pour l'analyse.
107. En outre, il pourrait être utile d'étudier/inclure des données plus anciennes pour savoir si certaines espèces/sous-groupes d'un groupe/catégorie d'aliments ont tendance à présenter des limites plus élevées.

108. Dans certains cas, il peut être approprié d'effectuer une analyse des tendances temporelles. Dans ces cas, les données de plus de 10 ans doivent être examinées pour déterminer si les concentrations ont changé/changent avec le temps et cela pourrait être utilisé pour déterminer si un certain nombre d'années de données devrait être utilisé pour l'élaboration des LM afin de représenter les concentrations actuelles.
109. Si un Code d'usages a été établi et mis en œuvre, les données examinées doivent porter sur les années suivant la mise en œuvre de ce Code d'usages afin de refléter les bonnes pratiques agricoles et de fabrication, sauf si un pays indique que les bonnes pratiques équivalentes ou plus strictes (nationales ou régionales) avaient déjà été mises en œuvre avant l'établissement du Code, auquel cas ces données peuvent être incluses.
- Les éléments susceptibles d'indiquer si un Code d'usages a été mis en œuvre peuvent être les suivants:
- Une baisse constante des niveaux après une certaine année, et
 - Différences de limites entre pays voisins au sein d'une même région qui ne peuvent être expliquées par des facteurs géographiques.
 - Informations sur les pratiques de mise en œuvre fournies par le pays qui soumet les données
110. Il est souvent difficile de juger, à partir des ensembles de données, si un Code d'usages a été mis en œuvre ou non. De préférence, des informations sur la mise en œuvre d'un Code d'usages sont demandées dans l'appel de données, et aussi si le pays qui soumet ces données a déjà mis en place ses propres mécanismes de surveillance. Si le GTE exclut des données en raison de l'absence de mise en œuvre d'un Code d'usages, les exclusions et leur justification doivent être clairement documentées dans le document final présenté par le GTE à la session plénière du CCCF.

ANALYSE STATISTIQUE DES DONNÉES D'OCCURRENCE /TRAITEMENT D'ENSEMBLES DE DONNÉES POUR L'ÉLABORATION DES LM

Considérations générales

111. Les rubriques suivantes expliquent les considérations à prendre en compte avant de procéder à une analyse statistique des données extraites/nettoyées, et la manière dont les résultats de l'analyse statistique doivent être présentés dans les GTE en vue de l'élaboration de LM applicables à l'échelle mondiale.
112. Les données extraites/nettoyées doivent faire l'objet d'une analyse statistique. Lors de l'évaluation des données, la première étape devrait consister à analyser la distribution de l'ensemble des données. En général, la distribution des données relatives aux contaminants dans les aliments tend à être asymétrique à droite, par exemple une distribution log-normale (voir Figure 1). Pour ces distributions non normales, l'utilisation de méthodes statistiques paramétriques, basées sur la distribution normale, n'est pas appropriée.
113. La NGCTAHA stipule à l'Annexe I que «les LM devraient être établies à un niveau (légèrement) supérieur au champ normal de variation des concentrations dans les produits de consommation humaine et animale». Cela signifie que pour élaborer une LM, il est nécessaire d'estimer des valeurs de centile élevées (généralement des valeurs du 95^e centile) avec un niveau de confiance significativement élevé. En matière de sécurité alimentaire, un niveau de confiance de 95 % est généralement utilisé. La figure ci-dessous (Figure 1) explique, à l'aide d'une distribution modélisée, la relation entre une valeur de centile élevée, une LM hypothétique (généralement une valeur arrondie de la valeur de centile) et le pourcentage de points de données qui dépassent la LM proposée lorsque la LM est identique à la valeur du 98^e centile.



EN	FR
Frequency of sample	Fréquence de l'échantillon
Hypothetical ML	LM hypothétique
Rejection rate (2%)	Taux de rejet (2 %)
Rejected from the market	Rejeté du marché
98 th percentile	98e centile
Percentile value (of concentration)	Valeur du centile (de la concentration)

Figure 1. Représentation simplifiée de la relation entre une valeur de centile élevée et une LM hypothétique, le taux de rejet et le pourcentage d'échantillons dépassant la LM proposée.

Remarque: Dans ce qui précède, on suppose que la LM hypothétique est identique à la valeur du 98^e centile. Le choix du 98^e centile pour les LM hypothétiques dans cette figure n'est donné qu'à titre d'exemple.

Nombre minimum de points de données pour l'estimation des valeurs de centile élevées

114. Pour élaborer une LM, il est nécessaire d'estimer les valeurs de centile élevées (généralement les valeurs du 95^e centile) d'un ensemble de données. Un nombre minimum de 59 points de données est requis pour une estimation du 95^e centile avec un niveau de confiance de 95 % (voir l'option 1, Annexe II). L'Annexe II fournit des détails supplémentaires sur les options alternatives pour le calcul du nombre minimum de points de données requis pour l'estimation des valeurs de centile élevées.

Traitement des ensembles de données

Traitement des ensembles de données avec un nombre de points de données faible

115. Lorsque l'évaluation du JECFA suggère qu'un risque sanitaire lié à l'exposition à un contaminant est important, un nombre de points de données inférieur au nombre minimum de points de données (c'est-à-dire 59 points de données, voir le paragraphe 114 et l'Annexe II) serait considéré comme adéquat pour élaborer une LM, à condition que le niveau de confiance des valeurs de centile élevées estimées ne soit que légèrement inférieur au niveau de confiance élevé attendu, tel que 95 %. (Voir l'Annexe II pour un exemple de calcul du niveau de confiance) Par exemple, lorsqu'une LM est nécessaire de toute urgence pour protéger la santé des consommateurs, le GTE chargé de recommander des LM devrait envisager de recommander au CCCF l'élaboration de LM, même si seul un petit nombre de points de données est disponible. Si des données suffisantes sont disponibles à l'avenir, une révision des LM précédemment établies peut être envisagée.
116. S'il n'y a pas de risque sanitaire immédiat et que le nombre de points de données est insuffisant pour élaborer une LM, des appels de données supplémentaires peuvent être demandés. Toutefois, si, après plusieurs appels de données, le nombre de points de données disponibles reste très inférieur au nombre minimum requis, il convient de recommander au CCCF, au cas par cas, de décider d'élaborer une LM en utilisant l'ensemble de données limité ou d'interrompre les travaux. Une autre option peut être d'étendre la LM à un groupe alimentaire plus important, si une LM pour le groupe plus important est justifiée.
117. Pour les produits de base qui ne sont pas couramment consommés et/ou qui ne font pas l'objet d'un commerce international, la disponibilité des données d'occurrence peut être insuffisante. Dans de tels cas, le GTE chargé d'établir les LM devrait envisager de recommander au CCCF que la LM demandée pour la combinaison produit de base/contaminant ne réponde pas aux critères décrits dans la NGCTAHA et le Manuel de procédure (Section 4.5 «Politique du Comité du Codex sur les contaminants dans les aliments en matière d'évaluation de l'exposition aux contaminants et aux toxines présents dans les groupes de denrées alimentaires ou d'aliments pour animaux») pour l'élaboration des LM.
118. Si le nombre de points de données est nettement inférieur au nombre minimal requis de points de données (c'est-à-dire 59, voir le paragraphe 114 et l'Annexe II) et qu'il n'y a pas de raison sérieuse d'élaborer immédiatement une LM, il n'est pas nécessaire d'effectuer d'autres analyses statistiques. Si des données supplémentaires sont nécessaires pour établir une LM statistiquement robuste, d'autres appels de données peuvent s'avérer nécessaires.
119. Lors de l'examen des LM existantes, même si seul un petit nombre de points de données provenant de régions limitées est disponible et/ou qu'aucune nouvelle donnée ne sera générée, le GTE chargé de recommander les LM ne devrait pas automatiquement recommander la révocation de la LM en raison du petit nombre de points de données, à moins que la LM ne soit incompatible avec les bonnes pratiques agricoles et de fabrication actuelles ou avec les données toxicologiques actuelles. S'il existe un risque potentiellement important lié à la consommation du produit de base, une option consisterait de recommander au CCCF de maintenir la LM

existante, ou s'il n'y a plus de risque important pour la santé ou de barrière commerciale connue, une option consisterait de recommander au CCCF de révoquer la LM. Dans certains cas, lors de la révision d'une LM, s'il n'y a qu'un petit nombre de points de données pour un produit de base particulier, il peut être possible d'envisager de fusionner le produit avec le groupe d'aliments dont le produit a été exclu à l'origine (par exemple, en supprimant l'exclusion des conserves de *Brassica* d'une LM sur les conserves de légumes (réf. REP18/CF paragraphe 32)).

Traitement des ensembles de données lorsque les données disponibles sur les produits de base/aliments individuels sont insuffisantes, mais que les données pour le groupe d'aliments sont suffisantes

120. Même lorsque les points de données sont suffisants pour un groupe d'aliments entier, si les données sont séparées en fonction des aliments individuels dans ce groupe d'aliments, le nombre de points de données peut être trop faible pour l'élaboration de LM pour les aliments individuels. En général, l'analyse des données actuellement disponibles devrait commencer par l'élaboration d'un document de travail. Sur la base des données disponibles, le GTE devrait recommander une approche préliminaire de l'établissement des LM pour les aliments individuels et les groupes d'aliments, et recommander une formulation pour un nouvel appel de données. Si, après l'appel et la collecte de données, il s'avère qu'il y a moins de points de données disponibles qu'initialement prévu, il peut être nécessaire de modifier l'aliment ou les aliments ciblé(s) par la LM afin d'élargir l'éventail des aliments, par exemple en remplaçant un aliment par un sous-groupe d'aliments ou des sous-groupes d'aliments par des groupes d'aliments.
121. La pertinence de l'élaboration d'une LM pour un groupe d'aliments dépend de la similitude de la distribution des valeurs de concentration de contaminants dans les aliments individuels au sein du groupe d'aliments. Des tests statistiques non paramétriques, tels que le test U de Mann-Whitney (pour 2 ensembles de données) ou le test H de Kruskal-Wallis (pour 2 ensembles de données ou plus), peuvent être utilisés pour déterminer si la distribution des aliments du groupe peut être considérée comme similaire, même lorsque le nombre de points de données est relativement faible. Toutefois, comme les taux de contamination entre les produits d'un même groupe alimentaire peuvent varier de manière significative, il est important de faire preuve de souplesse dans le choix des méthodes et l'interprétation des tests statistiques. Si le nombre de points de données est relativement faible, il est également utile de comparer les ensembles de données à l'aide de diagrammes en boîte à moustaches, à condition que le pourcentage de données censurées à gauche soit inférieur à 25 % de l'ensemble de données concerné.
122. Si un aliment individuel présente une distribution des concentrations de contaminants différente de celle des autres aliments du groupe, il peut être nécessaire d'élaborer deux LM différentes, l'une pour le groupe d'aliments à l'exclusion de l'aliment individuel et l'autre pour l'aliment individuel. Des approches similaires peuvent être prises pour le(s) sous-groupe(s) du groupe d'aliments. Si les données relatives à une ou plusieurs denrées alimentaires sont insuffisantes pour atteindre le nombre minimal de points de données requis, des appels de données supplémentaires peuvent être lancés pour les aliments pour lesquels il est jugé nécessaire d'élaborer des LM. Si la consommation d'un aliment individuel dont la distribution diffère de celle du groupe d'aliments ne contribue pas de manière significative à l'exposition totale au contaminant en question, elle peut être considérée comme négligeable du point de vue de la protection de la santé du consommateur. Dans de tels cas, aucun appel de données supplémentaire n'est requis et son exclusion de l'élaboration d'une LM pour le groupe de denrées alimentaires peut être envisagée (par exemple, LM pour le plomb dans le sel, qualité alimentaire excluant le sel des marais (REP18/CF paragraphes 39-41). En ce qui concerne les groupes d'aliments et leurs sous-groupes, il est possible de se référer aux produits couverts par les normes de produits Codex pertinentes, à la classification des denrées alimentaires et des aliments pour animaux du Comité du Codex sur les résidus de pesticides (CCPR) et à d'autres systèmes de catégorisation des aliments utilisés par le Comité du Codex sur les additifs alimentaires (CCFA) pour les denrées alimentaires transformées.
123. Lors de l'élaboration d'une LM pour un groupe d'aliments plus large en raison de la disponibilité limitée de données pour des aliments ou sous-groupes individuels, certains aliments (ou sous-groupes) peuvent présenter des schémas de distribution différents de ceux d'autres aliments du même groupe d'aliments. Si les données ne sont pas suffisantes pour élaborer une LM distincte pour ces aliments, le GTE chargé de recommander des LM pourrait recommander au CCCF d'exclure ces aliments de l'élaboration de la LM jusqu'à ce que des données suffisantes soient disponibles.

Traitement d'ensembles de données (y compris l'utilisation de méthodes de substitution) comportant une grande proportion de points de données censurés à gauche

124. Dans cette section, le terme «ensemble de données» fait référence à un ou plusieurs ensembles de données qui font partie des ensembles de données sélectionnés pour être utilisés dans le cadre de l'élaboration d'une LM. Cette section est particulièrement pertinente lorsque les ensembles de données utilisés pour l'élaboration des

LM contiennent un taux élevé de points de données non quantifiés (par exemple, en raison de la faible sensibilité des méthodes d'analyse disponibles pour la concentration dans les échantillons, d'une fréquence d'occurrence extrêmement faible, etc.

125. Bien qu'aucune définition officielle du terme «censuré à gauche» ne figure dans les documents du Codex, en statistique, les données individuelles sans valeurs quantifiées (finies) sont appelées «données censurées à gauche», généralement désignées comme des données inférieures aux limites de quantification (LOQ) et aux limites de détection (LOD) rapportées.
126. Pour l'analyse statistique d'ensembles de données contenant des données censurées à gauche, il convient d'envisager des méthodes de substitution conventionnelles, en particulier pour le calcul de statistiques telles que le 95^e centile, ou pour l'estimation des taux de rejet d'échantillons et des réductions d'exposition pour les produits de base cibles dans le cadre de LM hypothétiques. Si l'ensemble de données contient un taux élevé de données censurées à gauche, l'analyse statistique utilisant uniquement des valeurs quantifiées n'est pas recommandée, car cette pratique introduit un biais dans les résultats de l'analyse statistique.

Méthodes de substitution pour les ensembles de données comportant une grande proportion de données censurées à gauche

127. L'approche conventionnelle pour traiter les données censurées à gauche dans le cadre d'une analyse statistique consiste à utiliser un ou plusieurs des scénarios de substitution suivants:
 - Scénario de l'estimation basse: les résultats inférieurs à la LQ sont remplacés par zéro, ou par la LOD si celle-ci est connue (les résultats <LOD sont remplacés par zéro);
 - Scénario de l'estimation haute: les résultats inférieurs à la limite de quantification sont remplacés par la valeur de la limite de quantification rapportée; et
 - Scénario de l'estimation moyenne (MB): Estimation ponctuelle entre les deux scénarios extrêmes (estimation basse et haute); on attribue alors une valeur égale à la LOQ/2, à la racine carrée de la LOQ, ou à (LOD + LOQ)/2 si la LOD est connue, pour les résultats d'analyse inférieurs à la LOQ rapportée. En général, la LOQ/2 est la plus utilisée.

Pour chacun de ces scénarios, si la LOQ n'est pas rapportée et que seule la LOD l'est, utiliser la LOD comme alternative.

128. En général, en fonction de la distribution des données, ces méthodes de substitution peuvent être utilisées pour calculer des mesures de tendance centrale telles que la moyenne arithmétique lors du calcul de l'exposition alimentaire (voir Critères de santé environnementale 240 ou EHC 240¹¹). Si le GTE chargé de recommander des LM doit effectuer des calculs préliminaires pour la réduction de l'exposition, le choix des scénarios d'estimation basse, moyenne ou haute (LB, MB ou UB) peut affecter la moyenne arithmétique calculée et l'exposition estimée pour les produits cibles sur la base de la moyenne arithmétique. Toutefois, pour l'élaboration de la LM, l'effet des valeurs censurées à gauche sur l'estimation empirique des valeurs de centile élevées peut être négligeable et il peut y avoir peu d'impact sur la LM quel que soit le scénario choisi, à moins qu'une grande majorité des points de données soient censurés à gauche (c.-à-d. <LOD).
129. Les ensembles de données comportant une grande proportion de données censurées à gauche doivent être traités au cas par cas, en fonction de la toxicité du contaminant et de la consommation de l'aliment concerné. Idéalement, les éléments d'estimation basse, haute ou moyenne devraient être calculés et présentés. Il est très important de connaître la distribution des valeurs quantifiées en cas de pourcentage élevé de données censurées à gauche lors de l'estimation des valeurs de centile élevées à l'aide d'une distribution modélisée pour élaborer des LM.
130. Lorsque la dispersion des valeurs quantifiées se situe dans une fourchette étroite (valeurs proches les unes des autres) et qu'elle est proche de la LOQ rapportée, l'élaboration d'une LM pourrait ne pas être nécessaire, à moins que le contaminant ne soit hautement toxique. Le GTE peut faire une recommandation au CCCF sur l'opportunité d'une LM dans cette situation.

Traitement de plusieurs ensembles de données

Analyse des ensembles de données individuels et combinés, et prise de décision sur la nécessité de comparer les ensembles de données individuels avant de les combiner, en particulier lorsque les schémas de distribution sont significativement différents.

¹¹ ENVIRONMENTAL HEALTH CRITERIA 240, Principles and methods for the risk assessment of chemicals in food (WHO, 2009)

131. Comme les LM du Codex sont destinées à une application mondiale, elles devraient idéalement être basées sur des ensembles de données mondiaux. Bien que l'approche par défaut pour l'élaboration des LM consiste à utiliser l'ensemble des données mondiales combinées, les ensembles de données individuels par année ou par région sont fournis à titre d'information supplémentaire. Le CCCF doit décider au cas par cas si une LM doit être basée sur un ensemble de données mondiales ou sur un ensemble de données d'une région ou d'une année spécifique, après l'analyse statistique décrite dans cette section.
132. Des méthodes statistiques paramétriques sont disponibles pour comparer les schémas de distribution des ensembles de données individuels par région/pays ou par année. L'hypothèse nulle est que tous les ensembles de données sont supposés provenir de la même population. Ces tests comprennent le test U de Mann-Whitney (pour 2 ensembles de données) ou le test H de Kruskal-Wallis (pour 2 ensembles de données ou plus).
133. De nombreux modèles de méthodes statistiques non paramétriques sont disponibles sur Internet. Parmi eux, des modèles MS Excel pour effectuer les tests U de Mann-Whitney et les tests H de Kruskal-Wallis peuvent être téléchargés sur le site web de la JMPR de la FAO¹².
134. En outre, il est utile de tracer des diagrammes en boîte à moustaches ou des histogrammes pour chaque ensemble de données afin de comparer s'il existe des différences visuelles dans les distributions avant de combiner les ensembles de données. Il est préférable de ne dessiner un histogramme que lorsque l'ensemble de données contient un nombre suffisant de points de données (voir paragraphe 114 et Annexe II). Pour un ensemble de données avec un plus petit nombre de points de données, il est difficile de connaître la forme de la distribution à l'aide d'un histogramme, et un diagramme en boîte à moustaches est plus utile (voir Annexe III).
135. La proposition d'une ou plusieurs LM à l'aide d'un ou de plusieurs ensembles de données combinés (ensembles de données mondiales) a été faite de manière conventionnelle dans les GTE. Lorsqu'il n'y a pas de différence significative entre les distributions de plusieurs ensembles de données provenant de sources différentes, on considère qu'il est peu important d'effectuer des analyses et des comparaisons de données supplémentaires pour chaque ensemble de données individuel (bien qu'il serait idéal de le faire si les ressources et le temps le permettent).
136. Lorsque le nombre de points de données est très différent entre les ensembles de données individuels de différentes régions ou pays, l'ensemble de données combiné qui en résulte reflète principalement les conditions d'un pays ou d'une région ayant le plus grand nombre de points de données, plutôt que celles de tous les pays ou régions ayant fourni des données. Pour résoudre ce problème, il serait théoriquement possible, bien que cela nécessite un processus complexe, d'équilibrer les ensembles de données en les pondérant par le volume de production ou de commerce ou par tout autre facteur raisonnable. Toutefois, la méthodologie et la justification de l'utilisation de la pondération des données n'ont pas été examinées dans le passé par le CCCF et, en raison de sa complexité et de la charge de travail qui en découle, il n'est pas envisagé de le faire avant qu'une orientation pratique ne soit élaborée à l'avenir à cet effet. Si l'on craint une distorsion dans la distribution de l'ensemble de données combiné en raison de la présence d'un très grand ensemble de données, il convient d'examiner la question au cas par cas. Pour le moment, il peut être recommandé d'effectuer également des analyses pour chaque ensemble de données individuel, comme présenté dans les sections suivantes, ou de demander conseil au JECFA sur l'analyse des données.

Cas où l'analyse d'ensembles de données individuels est recommandée

137. Si un test statistique indique une différence significative entre la distribution de plusieurs ensembles de données, et que cette différence est importante, il est recommandé d'analyser les ensembles de données individuels en même temps que l'ensemble de données combiné pour l'élaboration de la LM. Toutefois, cette décision doit être prise au cas par cas, car l'ampleur des différences de distribution dépend généralement de la combinaison du produit de base et des contaminants examinés. Une justification de l'analyse des ensembles de données spécifiques séparément de l'ensemble de données combiné doit être fournie, et si aucune justification ne peut être trouvée, l'ensemble de données combiné doit être utilisé par défaut.
138. Lorsque l'on envisage d'utiliser des ensembles de données individuels, il est recommandé de comparer les résultats statistiques, tels que les valeurs de centile élevées des ensembles de données séparés, à ceux de l'ensemble de données combiné. Il convient de noter qu'il n'est pas possible d'obtenir des valeurs robustes de centile élevé pour les ensembles de données individuels ou combinés dont la taille des échantillons est inférieure au nombre minimum requis de points de données (voir paragraphe 114 et le Tableau 1 de l'Annexe II).
139. Il convient de noter que lorsque plusieurs ensembles de données sont examinés individuellement, plusieurs LM

¹² «Appendix XIV Electronic Attachments (2020_Nov)» et ouvrir «XIV 12 Spreadsheet for Kruskal_Wallis 20 group.xls» pour effectuer les tests U de Mann-Whitney et le test U de Kruskal-Wallis. <https://www.fao.org/agriculture/crops/thematic-sitemap/theme/pests/jmpr/jmpr-docs/en/>

possibles peuvent être identifiées. Il n'entre pas dans le champ d'application de la présente Orientation de fournir des conseils sur le choix éventuel d'une LM, car c'est le CCCF qui décide de la LM appropriée, en tenant compte d'éléments qui peuvent varier d'un cas à l'autre.

140. Si les ensembles de données provenant de différentes régions/pays sont analysés séparément au moyen des méthodes statistiques recommandées dans le présent guide, le GTE chargé de recommander les LM pourrait fournir des éléments pour le choix de l'ensemble de données sur lequel fonder la LM pour examen et décision par le CCCF. Par exemple, si l'on a l'assurance que les ensembles de données présentant des concentrations élevées concernent des produits fabriqués selon de bonnes pratiques (Code d'usages du Codex ou BPA, BPF, etc.), il pourrait être utile alors de se concentrer sur les ensembles de données présentant des concentrations élevées afin d'envisager des LM applicables à l'échelle mondiale.

Réalisation d'une analyse des données par visualisation

141. Il existe différentes méthodes pour illustrer les graphiques et les diagrammes afin de montrer la distribution des données d'occurrence et les valeurs statistiques permettant d'évaluer l'adéquation de l'avant-projet de LM. L'Annexe III présente des exemples d'analyses exploratoires de données et de statistiques. En fonction du contaminant, du nombre de points de données et de la distribution des données d'occurrence (paramétriques ou non), des analyses exploratoires des données par visualisation peuvent être effectuées au cas par cas.

Agrégation des données et calcul des statistiques descriptives

142. Les informations et les statistiques récapitulatives suivantes peuvent être présentées pour des ensembles de données d'occurrence répondant aux exigences minimales relatives:
- Nombre de points de données totaux.
 - Nombre de points de données inférieurs aux LOQ (ou LOD) rapportées, et/ou rapport entre le nombre de données <LOQ (ou < LOD) sur le nombre total de points de données.
 - Fourchette des LOQ (LOD) parmi les données exclues ou incluses dans l'ensemble de données final utilisé pour recommander les LM.
 - Moyenne (moyenne arithmétique): si l'ensemble de données contient des points de données inférieurs à la LOQ (LOD), trois moyennes arithmétiques basées sur trois scénarios d'estimation basse, moyenne ou haute (LB, MB ou UB) peuvent être préparées (si la distribution est normale ou proche de la normale et symétrique).
 - Si la distribution est très asymétrique, une moyenne géométrique utilisant la même approche que ci-dessus pourrait être une option, mais elle n'a pas encore été utilisée au sein du CCCF.
 - Médiane (valeurs du 50e centile), mais si plus de 50 % des points de données sont inférieurs à la <LOQ, la médiane peut être indiquée comme «<LOQ» (ou <LOQ);
 - Valeurs de centile élevées (par exemple, 95^e, 97^e et 98^e centiles, si nécessaire, en fonction des discussions au sein du GTE sur le(s) taux de rejet approprié(s)); si plus de 95 %, 97 %, etc. des points de données sont inférieurs à la LOQ, les centiles associés pourraient être indiqués comme «<LOQ» (ou LOD);
 - Minimum.
 - Maximum: dans les cas où le maximum a été identifié comme une valeur aberrante potentielle et que la valeur maximale n'a pas été exclue de l'ensemble de données, il peut être utile de déclarer la 2e valeur la plus élevée, la 3e valeur la plus élevée, etc. pour un contexte supplémentaire.
 - Plage de données quantifiées.
 - L'écart-type, qui est une mesure de la variation d'une distribution paramétrique.
 - Les valeurs interquartiles (voir Annexe III), qui sont une mesure de la variation d'une distribution non paramétrique.
143. Nombre de ces statistiques peuvent être facilement obtenues en utilisant les fonctions Excel, en utilisant un menu de statistiques descriptives dans les outils d'analyse de données de MS Excel, ou à partir de n'importe quelle autre application statistique. Les différentes applications statistiques utilisent des protocoles de calcul différents et renvoient donc des valeurs de centile différentes pour le même ensemble de données. Par conséquent, lors du calcul des valeurs de centile à l'aide d'applications informatiques, les valeurs obtenues doivent être soigneusement vérifiées par rapport aux fonctions utilisées et indiquer le nom et, le cas échéant, les détails de l'application utilisée pour le calcul.
144. Lorsque les données censurées à gauche constituent la majeure partie de l'ensemble de données, il peut être

impossible de calculer des valeurs de centile élevées. Dans de tels cas, il est recommandé d'utiliser la méthode de substitution avec les scénarios d'estimation basse, haute ou moyenne. Bien que cela doive être fait au cas par cas, en fonction du nombre de valeurs quantifiées et de la distribution des contaminants, des méthodes telles que l'estimation des valeurs de centile élevées à partir des fonctions de densité de probabilité en modélisant la distribution des données d'occurrence peuvent être utilisées.

Calcul des taux de rejet à des LM hypothétiques

Estimation des LM hypothétiques

145. À partir d'une valeur de centile élevée (généralement la valeur du 95^e centile) de l'ensemble de données cible pour l'analyse des données, on identifie une valeur candidate pour une LM, en tenant compte de la précision de la méthode d'analyse actuelle et des chiffres significatifs des résultats d'analyse. Les valeurs numériques utilisées pour les LM devraient, de préférence, être des chiffres arrondis en progression géométrique (0,01, 0,02, 0,05, 0,1, 0,2, 0,5, 1, 2, 5, etc.), à moins que cela ne crée des problèmes pour l'acceptabilité des LM (Annexe I de la norme CXS 193-1995)
146. Une fois que la valeur numérique candidate d'une LM a été déterminée, les valeurs supérieures et inférieures suivantes peuvent également être suggérées en tant que LM hypothétiques. (Par exemple, pour une LM candidate de 0,5 mg/kg, les LM hypothétiques supplémentaires pourraient être de 0,4 et 0,6 mg/kg). En cas de révision d'une LM existante, celle-ci doit également être ajoutée à la liste des LM hypothétiques. En outre, les valeurs obtenues à partir des valeurs de centile élevées (par exemple, les valeurs des 95^e, 97^e et 98^e centiles) peuvent également être utilisées directement comme LM hypothétiques.
147. Lorsqu'il est décidé d'analyser séparément plusieurs ensembles de données présentant des schémas de distribution différents, les LM hypothétiques sont déterminés à partir des valeurs de centile élevées de chaque ensemble de données. Si les schémas de distribution sont très différents, les LM hypothétiques des différents ensembles de données peuvent être très différents (voir paragraphe 139).
148. Il n'y a pas de règle concernant le nombre de LM hypothétiques à proposer, mais il est préférable d'avoir plus de 2 LM hypothétiques (à examiner au cas par cas), afin que l'impact sur l'exposition alimentaire aux produits de base cibles et les aspects économiques découlant des taux de rejet soient examinés plus en détail par le GTE chargé de recommander des LM au CCCF.
149. Les raisons du choix d'une LM doivent être clairement expliquées dans le document préparé pour le CCCF. La justification doit se fonder clairement sur l'analyse et non sur une valeur choisie à l'avance.

Calcul des taux de rejet à des LM hypothétiques

150. Le taux de rejet est défini par l'équation ci-dessous. Elle peut être facilement obtenue en utilisant directement les fonctions de MS Excel (telles que la fonction COUNTIF) ou en utilisant des applications statistiques ou de modélisation/simulation après avoir modélisé chaque ensemble de données. Si une méthode différente est utilisée pour calculer le taux de rejet, elle doit être clairement indiquée dans le document de travail.

Taux de rejet (%) = (nombre de points de données > LM hypothétique) / (nombre total de points de données) × 100
151. Il convient de noter que le taux de rejet peut être différent de celui prévu pour le centile élevé en raison de l'arrondissement. Plus le nombre de points de données utilisé pour calculer les taux de rejet est faible, plus l'incertitude dans l'estimation du taux de rejet est grande. Dans le calcul du taux de rejet, on suppose que les échantillons qui dépassent la LM hypothétique sont exclus du marché avec une probabilité de 100 % par l'application de la LM.

Évaluation de l'impact d'une LM sur le taux de rejet

152. Pour évaluer l'impact du produit de base sur le commerce international, il convient d'utiliser l'ensemble des données mondiales combinées et, si nécessaire, les ensembles de données pour chaque région. Il n'est pas recommandé que le GTE chargé de recommander les LM calcule les taux de rejet pays par pays, mais un membre du Codex pourrait attirer l'attention du GTE/CCCF sur l'impact économique de certaines LM hypothétiques pour son pays, au cas où cet impact ne serait pas suffisamment pris en compte dans l'évaluation de l'impact de la LM hypothétique sur le taux de rejet pour sa région ou son pays.
153. Pour les contaminants dont on sait que les concentrations varient fortement d'une année à l'autre, le taux de rejet pourrait être calculé pour l'ensemble des données de chaque année, si possible, afin de pouvoir comparer les taux de rejet d'une année à l'autre.

Amélioration du calcul des taux de rejet

154. Pour différentes LM hypothétiques, les taux de rejet calculés peuvent ne pas changer de manière significative, en fonction du modèle de distribution du contaminant. Étant donné la nature asymétrique des données sur les contaminants, le nombre de points de données dans la plage de centiles élevés est souvent beaucoup plus faible que dans la plage des centiles faibles, ce qui affecte l'estimation des LM hypothétiques et des taux de rejet (voir la Figure 1 pour la forme de la distribution).
155. Si le modèle de distribution de l'ensemble de données combiné (potentiellement mondial) présente un pic unique, une application de modélisation/simulation (telle que @Risk, Crystal Ball, R, etc.) peut être utilisée pour modéliser la distribution afin d'estimer en continu la distribution à proximité des valeurs de centile élevées de la fonction de distribution, et des estimations plus fines et améliorées du taux de rejet peuvent être possibles.
156. Si une évaluation d'impact plus détaillée ou améliorée concernant les taux de rejet est nécessaire, demander une évaluation du JECFA peut être une option.

Calcul préliminaire des effets des LM sur la réduction de l'exposition alimentaire au contaminant d'un produit de base cible à des LM hypothétiques

Calcul de l'exposition alimentaire et de son taux de réduction du produit de base cible à des LM hypothétiques

157. Pour s'assurer que la LM proposée protège la santé des consommateurs, il pourrait être opportun d'évaluer quantitativement l'effet d'une LM hypothétique sur la réduction de l'exposition alimentaire du produit de base cible en comparant l'exposition avec et sans LM. Dans le cas d'une révision d'une LM existante, l'exposition alimentaire du produit de base cible à la LM déjà établie est comparée à l'exposition aux nouvelles LM hypothétiques (LM révisées). Le calcul de l'exposition alimentaire étant une fonction d'évaluation des risques qui devrait être assumée par le JECFA, le CCCF peut demander au JECFA de procéder à une évaluation des effets des LM hypothétiques sur la réduction de l'exposition alimentaire si une évaluation détaillée est nécessaire. [Il est important de clarifier les rôles du JECFA et du CCCF en tant qu'évaluateur et gestionnaire des risques, respectivement, dans le calcul des taux de réduction de l'exposition alimentaire lors de l'examen des LM. Grâce à cette compréhension, une évaluation préliminaire de l'exposition au produit de base cible peut également être réalisée en tant qu'information de référence dans les GTE chargés de recommander des LM si les ressources sont disponibles. Pour des exemples de calcul du taux de réduction de l'exposition alimentaire du produit cible, se référer à l'Annexe IV]

Amélioration du calcul des taux de réduction de l'exposition alimentaire

158. Si une évaluation d'impact plus détaillée ou améliorée concernant l'exposition alimentaire est nécessaire, une évaluation du JECFA peut être demandée.

Préparation des recommandations finales au CCCF

159. Lors de la finalisation d'une analyse, le GTE chargé de recommander des LM peut être confronté à des questions difficiles telles que la nécessité d'une LM sur la base d'une évaluation préliminaire de l'exposition, la nécessité de plusieurs projets de LM proposés, l'ensemble de données à utiliser ou le taux de rejet approprié. Des perspectives différentes de celles envisagées lors de l'élaboration des TdR pour les travaux peuvent apparaître. Le GTE ne doit pas prendre de décision finale, mais préparer des questions et des recommandations à soumettre au CCCF dans son ensemble.
160. Si l'exposition alimentaire estimée dans le cadre de l'estimation supérieure est bien inférieure à la valeur indicative basée sur la santé (VABS), même en l'absence de LM, et qu'une LM proposée est égale ou proche de la valeur LOQ, la LM n'aurait qu'un faible impact sur la réduction de l'exposition alimentaire et ne serait pas nécessaire. Pour les contaminants pour lesquels aucune VBAS n'a été établie, si toutes les données sont censurées à gauche, il pourrait être recommandé au CCCF d'établir la ou les LM à la valeur de la LOQ actuelle, s'il existe un risque potentiel pour la santé. Toutefois, si la plupart des données sont censurées à gauche et qu'il n'y a pas ou peu de problèmes de santé, il convient de recommander au CCCF qu'il n'est pas nécessaire d'élaborer une LM. Par exemple, un ensemble de données combinées sur le plomb dans les œufs de poule frais après le nettoyage des données contenait 99 % de points de données censurés à gauche, allant de 0,001 à 0,257 mg/kg. La réduction de l'exposition calculée à des LM hypothétiques était faible, et la LM proposée se situait dans la fourchette des LOQ rapportées. Par conséquent, l'élaboration d'une LM pour le plomb dans les œufs de poule a été interrompue (réf. CX/CF 22/15/7).
161. Si l'exposition alimentaire d'estimation supérieure est proche ou supérieure à la VABS ou si la marge d'exposition est faible, l'élaboration d'une LM pourrait être recommandée au CCCF, même si une LM proposée est proche des LOQ rapportées, à condition qu'il existe une (des) méthode(s) analytique(s) validée(s) avec une LOQ appropriée. Si nécessaire, il peut être recommandé de faire des appels de données supplémentaires en utilisant des méthodes d'analyse plus sensibles (LOQ plus faibles).

162. Lorsque les valeurs quantifiées présentent une variation importante et atteignent des valeurs significativement élevées, il pourrait être recommandé au CCCF d'élaborer une LM afin d'éliminer les denrées alimentaires hautement contaminées du marché international. Si le contaminant est hautement toxique ou génotoxique/cancérogène et qu'il se trouve dans des aliments consommés en grandes quantités, il pourrait être recommandé au CCCF d'établir une LM pour protéger la santé des consommateurs, même si le taux de rejet est faible. Par exemple, l'ensemble de données combinées sur les aflatoxines totales dans le sorgho en grains après nettoyage contenait 94 % de données censurées à gauche, et la fourchette supérieure des concentrations quantifiées dans cet ensemble de données dépassait 200 µg/kg. Cela indique qu'une LM basée sur un faible taux de rejet aurait encore un impact important sur la réduction de l'exposition alimentaire aux aflatoxines du sorgho en grains. Un projet de LM a été recommandé au CCCF pour les aflatoxines dans le sorgho avec un taux de rejet de 0,9 % et une réduction de l'ingestion de 46,5 % (réf. CX/CF 22/15/9).
163. Globalement, si l'exposition alimentaire est évaluée, les GTE chargés de recommander des LM au CCCF devraient évaluer l'équilibre entre le taux de rejet et la réduction de l'exposition alimentaire à partir du produit cible pour chaque LM hypothétique, et déterminer le taux le plus bas qu'il est raisonnablement possible d'atteindre ou proposer des options au CCCF pour éclairer la décision du Comité.
164. Bien qu'il n'entre pas dans le champ d'application de cette Orientation de déterminer quel taux de rejet est le plus approprié, le GTE chargé de recommander l'élaboration des LM au CCCF, devrait également tenir compte des modèles de consommation régionaux et internationaux pour déterminer un projet de LM parmi les LM hypothétiques en ce qui concerne la protection de la santé des consommateurs, la garantie de la sécurité alimentaire et d'un commerce équitable.
165. Il incombe aux membres du Codex de vérifier l'impact du projet de LM (ou de LM hypothétique(s)) par rapport à leurs propres données nationales/régionales et de fournir des commentaires sur les résultats de leur analyse statistique au GTE ou à la session plénière.

PRESENTATION DES DONNEES DANS LES RAPPORTS DES GTE AU CCCF

Présentation de l'analyse des données: analyse statistique

166. Il est important que les données soient présentées de telle manière dans le rapport du GTE au CCCF en tant que document de travail de manière à permettre une discussion éclairée sur les LM appropriées à délibérer dans le cadre de la procédure par étapes du Codex. Cela signifie que les données seront rapportées en incluant toutes les hypothèses, par exemple le nombre de données exclues et les raisons de cette exclusion, la manière dont les données censurées à gauche ont été gérées, si les données en dehors de la base de données GEMS/Aliments ont été prises en compte, etc.
167. L'Annexe V fournit les éléments et les exemples de modèles qui peuvent être utilisés par le GTE lors de la présentation des résultats de l'analyse statistique des données d'occurrence pour l'élaboration de LM. Le GTE peut utiliser les modèles ou les modifier au cas par cas, car le détail rapporté dépendra de la quantité de données disponibles et de la nature du contaminant.

ANNEXE I

Détermination et traitement des valeurs aberrantes/extrêmes

Le terme «valeurs aberrantes» est défini dans le document du Codex (CXG 72-2009) comme suit:

Valeurs aberrantes: Un membre d'un ensemble de valeurs qui est incompatible avec d'autres membres de cet ensemble.

Remarque:

La pratique suivante est recommandée pour traiter les valeurs aberrantes.

a) Des tests tels que ceux de Cochran (pour les variations à l'intérieur d'un même laboratoire) ou de Grubb (pour les variations entre laboratoires) sont appliqués pour identifier les valeurs suspectes ou les valeurs aberrantes:

- si la statistique du test est inférieure ou égale à sa valeur critique de 5 %, l'élément testé est accepté comme correct;

- si la statistique du test est supérieure à sa valeur critique de 5 % et inférieure ou égale à sa valeur critique de 1 %, l'élément testé est appelé valeur suspecte et est indiqué par un astérisque simple;

- si la statistique du test est supérieure à sa valeur critique de 1 %, l'élément est appelé une valeur statistique aberrante et est indiqué par un astérisque double.

b) On examine ensuite si les valeurs suspectes et/ou les valeurs statistiques aberrantes peuvent s'expliquer par une erreur technique, par exemple:

- une erreur dans l'exécution de la mesure,

- une erreur de calcul,

- une simple erreur d'écriture lors de la transcription d'un résultat de test,

- l'analyse d'un mauvais échantillon.

S'il s'agit d'une erreur de calcul ou de transcription, le résultat suspect doit être remplacé par la valeur correcte; s'il s'agit d'une erreur d'analyse d'un mauvais échantillon, le résultat doit être placé dans la cellule correcte. Une fois cette correction effectuée, il convient de répéter l'examen pour détecter les valeurs suspectes ou les valeurs aberrantes. Si l'explication de l'erreur technique est telle qu'il s'avère impossible de remplacer le résultat de test suspect, il faut alors l'écarter comme une «véritable» valeur aberrante qui n'appartient pas à l'expérience proprement dite.

c) Lorsqu'il subsiste des valeurs suspectes et/ou des valeurs statistiques aberrantes qui n'ont pas été expliquées ou rejetées comme appartenant à un laboratoire atypique, les valeurs suspectes sont conservées comme éléments corrects, et les valeurs statistiques aberrantes sont éliminées, à moins que le statisticien ne décide, pour une raison valable, de les conserver.

Les valeurs aberrantes sont un type de valeurs extrêmes, généralement déterminées par des tests statistiques tels que décrits ci-dessus.

Il existe des tests statistiques pour déterminer les valeurs aberrantes (tels que les tests de Cochran ou de Grubb), mais ils supposent généralement une distribution normale. Par conséquent, elles ne conviennent pas à l'application des données sur la présence de contaminants dans les aliments si elles ne suivent pas une distribution normale ou une autre distribution qui peut être convertie en distribution normale.

Il est fait référence aux paragraphes 82-92 du document principal

Test statistique des valeurs aberrantes

L'approche interquartile (IQR) est utile pour identifier les valeurs aberrantes, car elle peut être appliquée à des données nonparamétriques. Cette méthode détermine les valeurs supérieures aux «valeurs du 75^e centile + 1,5 × écart interquartile» dans l'ensemble de données comme étant des valeurs aberrantes. L'approche IQR est largement utilisée comme méthode facile dans une variété de domaines pour identifier les valeurs aberrantes. L'approche IQR suppose que les points de données équivalents à la médiane $\pm 2,7 \sigma$ dans une distribution normale se situent dans l'intervalle de la médiane ± 2 IQR. Si la distribution est caractérisée par un aplatissement et une asymétrie élevés, comme dans le cas des données sur la présence de contaminants dans les aliments, de nombreuses données du côté droit (concentrations

plus élevées) peuvent être considérées comme des valeurs aberrantes. En effet, un coefficient d'aplatissement élevé se traduit par un IQR plus petit et un coefficient d'asymétrie élevé par une plus grande variabilité.

Étant donné que l'exclusion d'un grand nombre de points de données dans la plage de concentration supérieure d'un ensemble de données affectera de manière significative le calcul des valeurs de centile élevées en fonction du nombre total de points de données de l'ensemble de données et, par la suite, de toute proposition de LM, il n'est pas recommandé d'exclure des données en tant que valeurs aberrantes uniquement sur la base des résultats de l'approche IQR sans tenir compte du modèle de contamination (par exemple, homogène ou hétérogène).

Autres méthodes pour identifier d'éventuelles valeurs aberrantes

Une autre façon d'identifier d'éventuelles valeurs aberrantes consiste à inspecter visuellement les données à l'aide d'une distribution de fréquences et à identifier les données qui semblent déconnectées du reste des données. Cependant, ce n'est pas une base suffisante pour exclure les données déconnectées en tant que valeurs aberrantes. La figure 2 est un exemple de données de l'UE sur la somme des toxines T-2 et HT-2 dans les produits de mouture de l'avoine, où un nombre relativement faible de points de données présentent des niveaux supérieurs à 500 µg/kg. Dans de telles circonstances, et connaissant la nature du contaminant, à moins qu'il puisse être déterminé que ces niveaux sont définitivement en dehors de la variabilité naturelle du contaminant, ces données doivent rester dans l'ensemble de données.

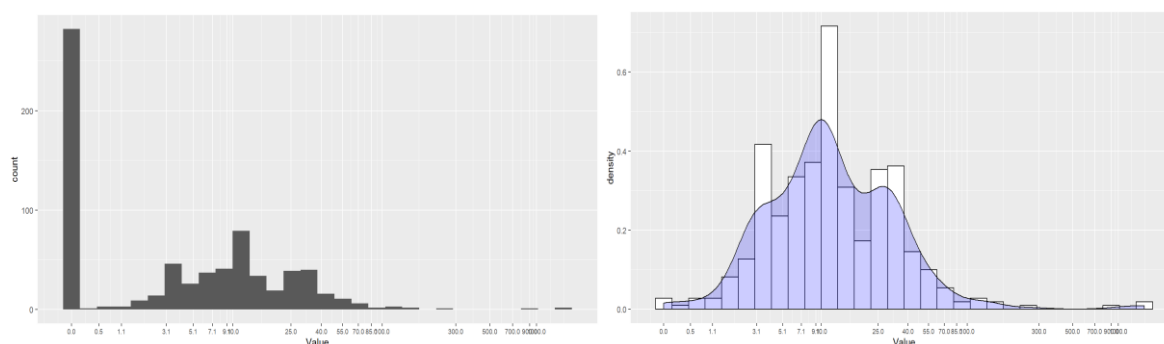


Figure 2. Exemple de données de l'UE sur la somme des toxines T-2 et HT-2 dans les produits de mouture de l'avoine (717 résultats dont 438 résultats quantifiés), (à gauche) Histogramme: tous les résultats, (à droite) Histogramme/courbe de densité de probabilité des résultats quantifiés

Décision sur le traitement d'éventuelles valeurs aberrantes

Dans certains cas, les valeurs extrêmes sont scientifiquement valables en fonction des conditions de production et météorologiques, et d'autres facteurs potentiels tels que les éruptions volcaniques, etc. Compte tenu des caractéristiques du modèle de distribution des données relatives à la présence d'un contaminant dans les aliments, il n'est pas recommandé d'exclure simplement les valeurs extrêmes sur la base des résultats des tests statistiques sur les valeurs aberrantes ou d'autres méthodes telles que l'inspection visuelle. Étant donné que la gamme des concentrations et des schémas de distribution qui peuvent être empiriquement ou théoriquement supposés varie considérablement en fonction du type de contaminant (métaux lourds, mycotoxines, etc.), le traitement des valeurs extrêmes doit être déterminé au cas par cas. Par exemple, une attention particulière doit être accordée aux mycotoxines dont les concentrations peuvent varier de manière significative en fonction des méthodes d'échantillonnage utilisées, en raison de la distribution hétérogène bien connue dans un lot, ainsi que de la variation annuelle très importante.

Les directives CXG 72-2009 supposent un ensemble de données de résultats d'analyses répétées du même échantillon qui présente une distribution normale. Elles stipulent que «les valeurs statistiques aberrantes sont éliminées à moins que le statisticien ne décide, pour de bonnes raisons, de les conserver». En revanche, les ensembles de données traités dans les présentes lignes directrices sont des résultats d'analyse provenant d'une variété d'échantillons et de différentes méthodes d'analyse. Comme on ne sait pas quelle distribution ils prendront et qu'ils peuvent être combinés à partir de sources multiples, il est difficile de prévoir l'étendue de la variation au sein d'un ensemble de données. Par conséquent, cette Orientation recommande que «les valeurs statistiques aberrantes ne soient pas écartées à moins qu'une bonne raison de les exclure ne soit identifiée et expliquée scientifiquement».

Si les valeurs extrêmes doivent être exclues en tant que valeurs aberrantes, il est recommandé que les raisons de l'exclusion soient clairement indiquées et que des analyses de sensibilité soient réalisées pour montrer comment l'exclusion ou la non-exclusion des valeurs aberrantes peut ou non affecter le calcul des valeurs de centile élevées. Il convient de rappeler que quelques valeurs extrêmes restant dans l'ensemble de données n'auront que peu d'effet sur le calcul des valeurs de centile élevé, à condition que le nombre total de points de données dans l'ensemble de données soit suffisamment supérieur au nombre minimum de points de données requis pour calculer les valeurs de centile élevé.

Exemples d'étude d'un ensemble de valeurs incompatibles avec d'autres membres de cet ensemble de données en tant que valeurs aberrantes possibles.

Nettoyage des données

- a) *Si la personne qui soumet les données confirme que les valeurs extrêmes sont dues à des erreurs, décider d'exclure ces données de toute analyse ultérieure:*
- *Valeurs aberrantes (clairement) dues à une falsification/action frauduleuse ou à une erreur humaine (par exemple, saisie incorrecte des données) → il peut être décidé, en consultation et en accord avec la personne qui soumet les données, d'exclure ces données de toute analyse ultérieure (ou de les noter comme «données invalides» après confirmation de cette personne)*
 - *aucune justification valable pour exclure ces données ne peut être fournie par la personne qui soumet les données ou peut être expliquée par le GTE pour ces éventuelles valeurs aberrantes → il peut être décidé de ne pas exclure ces données pour une analyse plus approfondie*
 - *une justification valable peut être fournie pour d'éventuelles valeurs aberrantes (telles que les données d'une année marquée par des conditions météorologiques extrêmes, les données d'une région ou d'un continent spécifique, etc.) → ces données ne doivent en principe PAS être exclues*

Analyse statistique

- b) *Évaluer l'impact d'éventuelles valeurs aberrantes sur les statistiques récapitulatives (moyenne arithmétique, valeurs de centile élevées).*

Les valeurs aberrantes éventuelles doivent être conservées dans l'ensemble de données, à moins qu'une analyse de sensibilité n'ait un impact significatif sur les statistiques récapitulatives. En cas de différence significative dans les résultats d'une analyse de sensibilité, le GTE ou le CCCF décidera au cas par cas si cette différence est significative, en fonction de la toxicité du contaminant et du type d'aliment, et donc s'il convient d'exclure ou non les valeurs aberrantes éventuelles.

- c) *Effectuer un test de valeurs aberrantes pour détecter les valeurs extrêmes qui nécessitent un examen plus approfondi. Les statisticiens doivent recommander des tests de valeurs aberrantes adaptés à l'ensemble de données, qui seront utilisés par le GTE du CCCF, le cas échéant. L'approche IQR peut être une option, mais l'exclusion automatique des valeurs aberrantes devrait être découragée à moins qu'il n'y ait une autre justification pour les exclure. En général, les tests de valeurs aberrantes qui nécessitent une distribution normale ne sont pas recommandés.*

ANNEXE II

Nombre minimum de points de données pour l'estimation des valeurs de centile élevées

Il est fait référence au paragraphe 54 du document principal qui mentionne que pour l'élaboration d'une LM, il est nécessaire d'estimer les valeurs du centile élevées (généralement les valeurs du 95^e centile) d'un ensemble de données. Un nombre minimum de 59 points de données d'échantillons est requis pour une estimation du 95^e centile avec un niveau de confiance de 95 %.

Cette Annexe II fournit des détails sur les différentes options alternatives pour le calcul du nombre minimum de points de données requis pour l'estimation des valeurs de centile élevées.

Trois options sont disponibles pour calculer le nombre minimum de points de données requis pour l'estimation des valeurs de centile élevées.

- **Option 1:** Calcul basé sur le concept qu'un ensemble de données contient une ou plusieurs valeurs supérieures à un certain centile se produisant avec une probabilité à un certain niveau de confiance. Cette méthode est basée sur la distribution binomiale. Le nombre minimum de points de données est obtenu à partir de la formule suivante:

$$n = \log(1-NC) / \log(p)$$
(NC=1- α , niveau de confiance (0<NC<1); p, centile/100 (0<p<1); et n, nombre de points de données).
 En général, un niveau de confiance de 95 % (NC=0,95) est utilisé dans le domaine de la sécurité alimentaire. Si un niveau de confiance plus élevé est nécessaire, n peut être calculé avec un niveau de confiance supérieur à 0,95.
- **Option 2:** Calcul basé sur la règle de Kroes *et al.* (2002)¹³. Le nombre minimum de points de données pour les valeurs de centile élevées (>75^e centile) peut être obtenu à partir de la formule suivante:

$$n \geq 8 / (1-p)$$
(p, centile/100 (0<p<1); et n, nombre de points de données).
 Aucune information n'est disponible dans la référence sur les niveaux de confiance.
- **Option 3:** Calcul basé sur les descriptions de Conover (1971)¹⁴ en utilisant une distribution binomiale et sur le concept qu'un ensemble de données contient une ou plusieurs valeurs supérieures à un certain centile et une ou plusieurs valeurs inférieures au centile se produisant avec une probabilité à un certain niveau de confiance. Le nombre minimum de points de données est obtenu à partir de la formule suivante:

$$n \approx 1/4 * x_{1-\alpha} * (1+p) / (1-p) + 1/2$$
(NC=1- α , niveau de confiance (0<CL<1); p, centile/100 (0<p<1); n, nombre de points de données; et $x_{1-\alpha}$, (1- α) quantile de la variable aléatoire χ^2 et valeur de $x_{1-\alpha}$ sont 9,488 ($\alpha=0,05$), 11,14 ($\alpha=0,025$) et 13,28 ($\alpha=0,01$))

La formule utilisant le même concept que l'option 1 (en remplaçant le centile/100 par (1-taux de violation/100)) a été utilisée pour les tests de conformité développés par le CCPR (CXG 33-1999) et le CCRVDF (CXG 71-2009). Certaines lignes directrices^{15,16} recommandent les chiffres obtenus à partir des options 2 et 3 (à un niveau de confiance de 99 %) pour un plan d'enquête visant à collecter des données sur la consommation alimentaire. Le tableau suivant montre le nombre minimum de points de données calculé pour obtenir les valeurs des 95^e, 96^e, 97^e, 97,5^e et 98^e centiles en utilisant les trois options ci-dessus à un niveau de confiance de 95 %.

Le Tableau 1 sert de guide pour comprendre le nombre minimum de points de données nécessaires pour élaborer des LM hypothétiques ou pour proposer des LM. Un tel ensemble de données est généralement/idéalement un ensemble de données global qui consiste en des ensembles de données individuels soumis à GEMS/Aliments ou directement au CCCF par des pays et/ou des organisations ou des observateurs, et qui couvre l'occurrence de la contamination dans le monde entier. Si les ensembles de données peuvent être combinés, il n'est pas nécessaire que chacun des ensembles de données individuels soumis par le(s) pays/organisation(s) membre(s) ou le(s) observateur(s) contienne plus de points de données que le nombre minimum indiqué dans le Tableau 1. Toutefois, lorsque des ensembles de données individuels par région et/ou par année sont utilisés séparément pour obtenir des valeurs de centile élevées plutôt que, ou en plus, de l'ensemble de données global, le nombre minimum de points de données est requis.

¹³ Kroes et al. 2002 "Assessment of intake from the diet." Food and Chemical Toxicology 40.2-3: 327-385.

¹⁴ Conover, 1971. Practical nonparametric statistics. Wiley, New York, États-Unis.

¹⁵ EFSA Journal, 2009. Principes généraux pour la collecte de données nationales sur la consommation alimentaire dans la perspective d'une enquête alimentaire paneuropéenne. 7(12):1435

¹⁶ EFSA Journal, 2014. Guidance on the EU Menu methodology. 12(12):394

Tableau 1. Nombre minimum de points de données pour obtenir des valeurs de centile élevées.

	Nombre minimum de points de données pour obtenir les valeurs de centile suivantes				
Option	95 ^e	96 ^e	97 ^e	97,5 ^e	98 ^e
Option 1 (NC=0,95)	59	74	99	119	149
[Option 1 (NC=0,99)]	[90]	[113]	[152]	[182]	[228]
Option 2 (NC: non fourni*)	160	200	267	320	400
Option 3 (NC=0,95)	93	117	157	188	236
[Option 3 (NC=0,99)]	[130]	[164]	[219]	[263]	[330]

* Aucune information sur le niveau de confiance n'est décrite dans la littérature originale de Kroes *et al.* (2002).

Le Tableau 1 et la formule pour chaque option peuvent être utilisés pour déterminer comment une valeur de centile élevée peut être calculée avec un niveau de confiance élevé (par exemple 95 %) à partir du nombre de points de données dans le nettoyage de l'ensemble de données, au cas par cas. Les chiffres dérivés de l'option 1 ont l'avantage de nécessiter le plus petit nombre de points de données par rapport aux autres options et sont les plus réalisables et les plus utilisés dans l'élaboration précédente de la LM par le CCCF en tenant compte du nombre de données disponibles pour les GTE.

Calcul du niveau de confiance

Le niveau de confiance pour l'option 1 du Tableau 1 peut être obtenu à l'aide de la formule suivante.

$$NC=1-p^n$$

(NC=1- α , niveau de confiance (0<NC<1); p, centile/100 (0<p<1); et n, nombre de points de données)

ANNEXE III

Génération de diagrammes/graphiques et de courbes se rapportant à la distribution des données d'occurrence

Comme première étape de l'analyse statistique, il est recommandé de créer des histogrammes ou des diagrammes en boîte à moustaches pour chaque ensemble de données (par exemple, les ensembles de données individuels et combinés) afin d'obtenir une perspective sur les tendances dans le modèle de distribution des données d'occurrence. Cette approche est appelée analyse exploratoire des données en statistique. Les histogrammes et les diagrammes en boîte à moustaches peuvent être créés à l'aide d'applications d'analyse statistique ou de tableurs, tels que Microsoft Excel.

Bien que diverses applications puissent être utilisées, Microsoft Excel offre une méthode simple. Pour effectuer des dessins et des analyses statistiques dans Microsoft Excel, il convient d'installer «Analysis ToolPak» à partir des modules complémentaires d'Excel afin d'utiliser diverses fonctions utiles pour l'analyse statistique. Un ruban «Analyse des données» sera ajouté à l'onglet «Données» de MS Excel après l'installation des compléments.

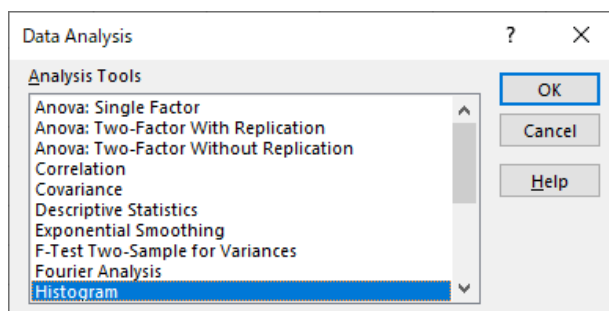


Figure 1. Exemple de menu de l'outil d'analyse des données dans Microsoft Excel

Les histogrammes peuvent être créés à partir du menu «Dessiner des graphiques» ainsi qu'à partir de l'outil Analyse de données de MS Excel. Toutefois, il est recommandé d'utiliser l'outil Analyse des données, qui offre une plus grande flexibilité dans la personnalisation du dessin du graphique. Les diagrammes en boîte peuvent être créés à partir du menu «Dessiner un graphique», et non à partir de l'outil «Analyse de données».

En général, les histogrammes donnent une bonne indication des schémas de distribution lorsque les points de données sont suffisamment nombreux (environ 50 ou plus). Le nombre minimum de points de données nécessaires pour calculer les valeurs des centiles élevées peut servir de guide pour le nombre approximatif de points de données nécessaires pour dessiner un histogramme. Si le nombre de points de données contenus dans les ensembles de données individuels diffère, l'axe vertical doit représenter la fréquence relative pour faciliter la comparaison.

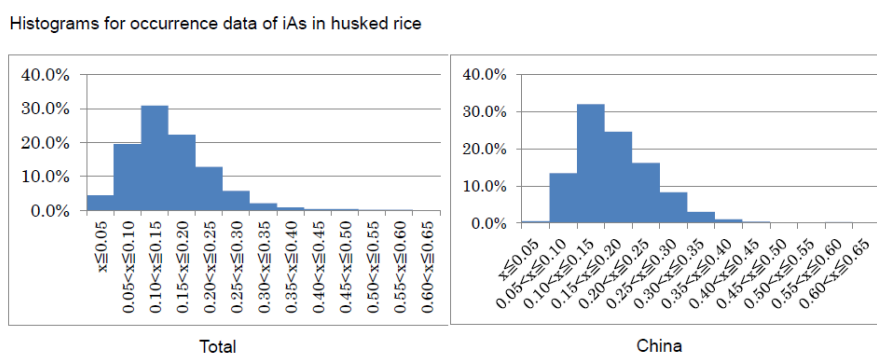


Figure 2. Exemple d'histogrammes de données sur la présence d'arsenic inorganique dans le riz décortiqué (ensemble de données combinées et individuelles) (réf. CX/CF 15/9/7)

Une courbe de fréquence cumulée peut également être ajoutée à un histogramme. Cependant, MS Excel ne peut à lui seul tracer des courbes de distribution. Il convient d'utiliser une application d'analyse statistique spécialisée (par exemple, SAS, SPSS, R, etc.) ou des applications complémentaires capables de modéliser/simuler (par exemple, @Risk, Crystal Ball, etc.).

Pour un ensemble de données comportant certains ensembles individuels avec un petit nombre de points de données à évaluer séparément, il est difficile de connaître la forme de la distribution à l'aide d'un histogramme pour ces ensembles de données individuels, et un diagramme en boîte à moustaches est plus utile, car il peut être créé même lorsque le nombre de points de données est inférieur à 20. Par exemple, les diagrammes en boîte à moustaches suivants pour les données de concentration d'arsenic inorganique dans le riz décortiqué provenant de différents pays peuvent

être tracés lorsque le nombre de données soumises par plusieurs pays est trop faible (par exemple, $n=9$) pour permettre de tracer un histogramme. Les diagrammes en boîte à moustaches ont été établis pour comparer les différents ensembles de données.

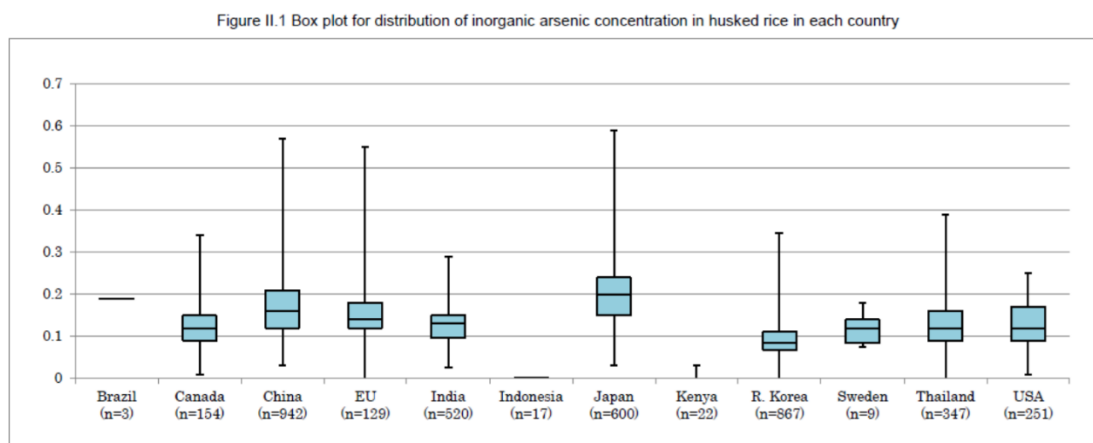


Figure 3. Les diagrammes en boîte à moustaches des ensembles de données individuels sur l'arsenic inorganique dans le riz décortiqué (réf. CX/CF 16/10/5)

Après avoir tracé des histogrammes ou des diagrammes en boîte à moustache, il est nécessaire de vérifier les éventuelles valeurs aberrantes et les différences dans les modèles de distribution des ensembles de données individuels, les formes des distributions, ainsi que la tendance centrale et l'étendue de l'ensemble de données. La présence de multimodalités dans l'ensemble des données combinées doit également être vérifiée. Lorsque l'ensemble de données combinées comporte des valeurs aberrantes ou présente clairement une distribution multimodale, il est nécessaire de reconsidérer la manière de traiter l'ensemble de données.

[Annexe IV]

Calcul préliminaire de l'exposition alimentaire et de son taux de réduction aux contaminants d'un ou des produits de base cibles aux LM hypothétiques

L'Annexe IV est mise entre crochets, car il est important de clarifier d'abord les rôles du JECFA et du CCCF en tant qu'évaluateur et gestionnaire des risques, respectivement, dans le calcul des taux de réduction de l'exposition alimentaire lors de l'examen des LM.

Pour tous les contaminants, l'exposition alimentaire à long terme/chronique peut être calculée à l'aide de l'équation suivante:

Exposition alimentaire ($\mu\text{g}/\text{personne.jour}$) = concentration ($\mu\text{g}/\text{g}$) $\times$ consommation alimentaire ($\text{g}/\text{personne.jour}$), ou

Exposition alimentaire ($\mu\text{g}/\text{kg pc.jour}$) = concentration ($\mu\text{g}/\text{g}$) $\times$ consommation alimentaire ($\text{g}/\text{kg pc.jour}$)

Il est fait référence au paragraphe 96 du document principal.

Pour estimer l'exposition alimentaire moyenne dans le cadre des LM hypothétiques, les concentrations moyennes arithmétiques (ou les concentrations moyennes géométriques dans le cas où la distribution est fortement asymétrique) dans le cadre des LM hypothétiques doivent être calculées à l'aide d'un ensemble de données dont sont exclues les données relatives à des concentrations supérieures à chaque LM hypothétique. Dans le calcul, on suppose que les échantillons dont les concentrations sont supérieures à la LM hypothétique sont exclus du marché avec une probabilité de 100 % par l'application de la LM.

Lorsque l'ensemble de données d'occurrence contient des données censurées à gauche et que la distribution le permet (par exemple, distribution normale), il est possible d'utiliser des moyennes arithmétiques calculées par les scénarios de substitution LB, MB ou UB. Pour l'évaluation de l'impact, il n'est pas nécessaire d'effectuer les calculs en utilisant les trois scénarios, mais le scénario utilisé pour calculer la moyenne arithmétique doit être indiqué.

Outre la moyenne arithmétique, la moyenne géométrique, la médiane ou les valeurs de centile élevées peuvent être utilisées au cas par cas, en particulier lorsque la distribution est fortement asymétrique, pour aider à comprendre clairement où se situe la tendance centrale à différentes fins. La valeur utilisée pour le calcul doit être clairement indiquée.

Si un ensemble de données contient des données de concentration beaucoup plus élevées que la LM hypothétique, la concentration moyenne sera considérablement réduite après l'exclusion de ces valeurs extrêmement élevées (voir la partie sur la détermination des valeurs aberrantes/extrêmes et leur traitement - paragraphes 82 -92). La médiane étant une statistique robuste, si le nombre de points de données dans l'ensemble de données ne change pas de manière significative après l'exclusion des points de données qui sont beaucoup plus élevés que la LM hypothétique, la médiane sous la LM hypothétique peut changer peu par rapport à la médiane sans LM.

Un modèle de calcul des estimations ponctuelles de l'exposition alimentaire chronique (Apport alimentaire estimé international, IEDI) à l'aide des régimes alimentaires par modules de consommation de GEMS/Aliments (expliqués ci-dessous) est disponible à l'adresse URL du programme GEMS/Aliments.¹⁷ Dans ce modèle, tous les pays sont rassemblés en 17 groupes et pour chaque groupe, il y a des données sur la consommation alimentaire dérivées des données du bilan alimentaire de la FAO et de quelques données supplémentaires fournies par les gouvernements. Après avoir saisi les données de concentration (dans ce cas, les concentrations moyennes arithmétiques calculées pour chaque LM hypothétique ou d'autres valeurs alternatives) et cliqué sur le bouton «Créer tableau», les apports moyens calculés pour 17 modules seront affichés.

Pour les aliments dont les données de consommation ne sont pas disponibles dans le régime alimentaire par module de consommation de GEMS/Aliments, il peut être possible d'estimer la consommation par habitant dans la population à partir des données disponibles, telles que les volumes de production. En ce qui concerne les données sur la consommation alimentaire, d'autres bases de données sont disponibles (par exemple, FAO/WHO GIFT, FAO/WHO CIFOCO, etc.). Les données utilisées et la manière dont elles ont été traitées devraient être indiquées dans le rapport du GTE.

Le pourcentage de diminution de l'exposition alimentaire au contaminant provenant des aliments ou groupes d'aliments concernés dans le cadre d'une LM hypothétique par rapport à l'exposition sans LM est considéré comme un taux de réduction de l'exposition.

Taux de réduction de l'exposition (%) = (exposition sans LM - exposition avec LM) / (exposition sans LM) $\times$ 100

Pour un contaminant pour lequel une ARfD a été établie par le JECFA, l'exposition aiguë/à court terme doit être calculée à l'aide du modèle de l'apport à court terme estimé international (IESTI) disponible à la même adresse URL que le modèle de l'IEDI. L'exposition aiguë/à court terme dans le cadre d'une LM hypothétique devrait être bien inférieure à

¹⁷ <https://www.who.int/teams/nutrition-and-food-safety/databases/global-environment-monitoring-system-food-contamination>

l'ARfD pour la population générale ou les enfants de 6 ans et moins, ou si l'ARfD est fixée pour les femmes en âge de procréer, elle devrait être bien inférieure à cette ArfD. Une ARfD n'a été recommandée que pour le DON et les composés apparentés; la manière de procéder au calcul de l'IESTI n'est pas expliquée en détail ici.

Si une valeur indicative basée sur la santé (VABS) est établie par le JECFA pour un contaminant/une toxine (PMTDI/PTDI, PTWI, PTMI ou ARfD, etc.), il peut être utile d'évaluer l'impact sur le pourcentage d'exposition provenant de la denrée alimentaire à laquelle la LM s'applique par rapport à la VABS. Il convient d'obtenir des informations sur le poids corporel moyen pour comparer l'exposition alimentaire à la VABS.

Rapport à la VABS (%) = exposition à la LM ($\mu\text{g}/\text{personne}/\text{jour}$) / poids corporel moyen ($\text{kg}/\text{personne}$) / VABS ($\mu\text{g}/\text{kg pc}/\text{jour}$) $\times 100$

Si les VABS sont fixées par semaine ou par mois, l'exposition doit être multipliée par le facteur approprié, par exemple 7 ou 30.]

ANNEXE V

Présentation de l'analyse des données/analyse statistique

(L'Annexe III pourrait être intégrée dans la présente Annexe étant donné qu'elle concerne la visualisation de la distribution des données d'occurrence, qui est un aspect important de la présentation de ces données).

Cette Annexe fournit les éléments et les exemples de modèles qui peuvent être utilisés par le GTE lors de la présentation des résultats de l'analyse statistique des données d'occurrence pour l'élaboration de la LM. Le GTE peut utiliser les modèles ou les modifier au cas par cas, car le détail du rapport dépend du nombre de points de données disponibles et de la nature du contaminant.

Pour chaque ensemble de données fourni par les membres (et les observateurs) et utilisé dans l'analyse statistique, les éléments suivants doivent être signalés:

- Nombre de points de données et proportion de données <LOQ.
- Moyenne arithmétique (LB, MB et/ou UB), médiane, minimum et maximum.
- Valeurs de centile pertinentes (par exemple, 95^e, 97^e, 98^e); et
- Graphiques ou diagrammes montrant des schémas de distribution (tels que des histogrammes ou des diagrammes en boîte à moustaches)

Le Tableau 1 peut servir de modèle pour la présentation des éléments ci-dessus.

Tableau 1. Exemple de modèle: Résumé des statistiques des ensembles de données

Aliments ou groupes d'aliments	Nombre total de points de données	Nombre ou pourcentage <LOQ	Moyenne (LB-UB) (mg/kg)	Médiane (LB-UB) (mg/kg)	95 ^e centile (LB-UB) (mg/kg)	97 ^e centile (LB-UB) (mg/kg)	98 ^e centile (LB-UB) (mg/kg)	Min. (mg/kg)	Max. (mg/kg)

Notes au Tableau 1: À remplir pour l'ensemble de données combiné (potentiellement l'ensemble de données global), centile: valeur de centile.

Si nécessaire, les éléments suivants peuvent également être utiles pour l'interprétation de l'ensemble des données:

- plage des LOQ rapportées; et
- écart-type (sans biais) le plus pertinent pour la moyenne LB et la moyenne UB.

Si la présence d'un aliment ou d'un groupe d'aliments varie considérablement d'une année à l'autre, il convient de fournir une analyse des données par année en utilisant la Tableau 2b et, en cas de différence statistiquement significative dans le modèle de distribution par région géographique, l'analyse devrait envisager de présenter les données par région (par exemple, continent ou région du Codex) à l'aide du Tableau 2a.

Tableau 2a. Exemple de modèle: Résumé des statistiques des ensembles de données par région pour un produit de base/aliment spécifique (catégorie)

Région/Continent/Monde	Nombre total de points de données	Nombre ou pourcentage <LOQ	Moyenne (LB-UB) (mg/kg)	Médiane (LB-UB) (mg/kg)	95 ^e centile (LB-UB) (mg/kg)	97 ^e centile (LB-UB) (mg/kg)	98 ^e centile (LB-UB) (mg/kg)	Min. (mg/kg)	Max. (mg/kg)

Notes au Tableau 2a: À remplir par région ou au niveau mondial, centile: valeur du centile.

Tableau 2b. Exemple de modèle: Résumé des statistiques des ensembles de données par année pour un produit de base/aliment spécifique (catégorie)

Année	Nombre total de points de données	Nombre ou pourcentage <LOQ	Moyenne (LB-UB) (mg/kg)	Médiane (LB-UB) (mg/kg)	95 ^e centile (LB-UB) (mg/kg)	97 ^e centile (LB-UB) (mg/kg)	98 ^e centile (LB-UB) (mg/kg)	Min. (mg/kg)	Max. (mg/kg)

Notes au Tableau 2b: À compléter par année/mondial ou par année/région, centile: valeurs de centile

Si le nombre de points de données est supérieur au nombre minimum de points de données requis pour un aliment donné ou si la distribution des concentrations pour un aliment donné diffère de façon significative de celle des autres aliments malgré le nombre inférieur de points de données, il est important de présenter un résumé des données pour tous les aliments individuels au sein d'un groupe d'aliments, ainsi qu'un résumé des données pour le groupe d'aliments dans son ensemble. Ce type d'analyse permet de comprendre l'impact d'une proposition de LM sur les différents aliments et de déterminer si l'élaboration d'une LM pour une grande catégorie d'aliments est plus appropriée que celle d'une LM pour des aliments individuels au sein de cette grande catégorie.

En général, deux à quatre LM hypothétiques sont estimées sur la base de la distribution. La présentation des données doit couvrir ces LM hypothétiques en montrant comment elles affectent les estimations de l'exposition alimentaire et quels seraient les taux de rejet à l'aide du Tableau 3.

Si nécessaire, un tableau, utilisant le Tableau 3 comme modèle, devrait être préparé non seulement pour l'ensemble de données combiné (potentiellement mondial), mais aussi pour les ensembles de données régionaux afin d'examiner l'impact des LM par zone géographique.

Tableau 3. Exemple de modèle: Résumé de l'analyse d'impact au niveau des LM hypothétiques

LM hypothétiques (mg/kg)	Concentration moyenne (mg/kg)	Apport alimentaire (µg/personne/jour)	Réduction de l'exposition (%)	Taux de rejet (%)
Nom de l'aliment ou du groupe d'aliments (nombre total de points de données)				

Notes au Tableau 3: Réduction de l'exposition = pourcentage de diminution de l'apport alimentaire du contaminant provenant de l'aliment ou du groupe d'aliments concerné

Enfin, un document de travail préparé pour être discuté lors de la session plénière du CCCF devrait être accompagné d'un tableau, en utilisant le Tableau 4 comme modèle et conformément au format de la NGCTAHA, qui comprend des informations sur l'avant-projet proposé de LM ainsi que des notes explicatives pour les produits de base examinés. L'explication de ce format et de son contenu est fournie dans la NGCTAHA.

Tableau 4. Format du tableau dans la NGCTAHA

Nom du produit de base/produit	Avant-projet de LM (mg/kg)	Portion du produit de base/produit à laquelle s'applique la LM	Notes/Remarques

ANNEXE VI

Glossaire des termes

Terme	Définition / Explication
Dose de référence aiguë (ARfD)	Estimation de la quantité d'une substance dans les aliments ou l'eau potable, exprimée sur la base du poids corporel, qui peut être ingérée dans une période de 24 heures ou moins sans risque appréciable pour la santé du consommateur. Elle est calculée sur la base de tous les faits connus au moment de l'évaluation. La dose de référence aiguë ARfD est exprimée en milligrammes de la substance chimique par kilogramme de poids corporel. (OMS, EHC 240)
Diagramme en boîte à moustaches	Méthode graphique permettant de présenter les caractéristiques importantes d'un ensemble d'observations. L'affichage est basé sur le résumé des données en cinq nombres, la partie «boîte» couvrant l'intervalle interquartile et les «moustaches» s'étendant pour inclure toutes les observations sauf les observations extérieures, celles-ci étant indiquées séparément. Elle est souvent particulièrement utile pour comparer les caractéristiques de différents échantillons. (Cambridge Dictionary of Statistics) Une boîte contient des données qui se situent entre le 25^e et le 75^e centile et les extrémités de la moustache indiquent en général les valeurs minimales et maximales. Les valeurs aberrantes sont tracées à l'écart de la moustache. La médiane (50^e centile) est indiquée dans l'encadré au-dessus du 25^e centile et au-dessous du 75^e centile. La moyenne est représentée par un symbole, tel que «X».
Test de Cochran	L'un des tests utilisés pour identifier les valeurs aberrantes, qui est un test des variabilités intra-laboratoire. Il faut d'abord l'appliquer, puis prendre les mesures nécessaires, en répétant les tests si nécessaire. (ISO 5725-2)
Niveau de confiance	Mesure de la fiabilité d'un résultat. Un niveau de confiance de 95 % ou 0,95 signifie qu'il existe une probabilité d'au moins 95 % que le résultat soit fiable.
Valeur critique	Valeur de la concentration ou de la quantité nette dont le dépassement conduit, pour une probabilité d'erreur α donnée, à décider que la concentration ou la quantité de l'analyte dans le matériau analysé est supérieure à celle du matériau vierge. Elle est définie comme suit: $\Pr(\hat{L} > L_c L=0) \leq \alpha$ Où : $\hat{L}$ est la valeur estimée, L est l'espérance ou la valeur réelle, et L_c est la valeur critique. (CXG72-2009)
Courbe de distribution	Graphique des fréquences des différentes valeurs d'une variable dans une distribution statistique.
Évaluation de l'exposition	L'évaluation de l'exposition est l'évaluation qualitative et/ou quantitative de l'absorption probable d'agents biologiques, chimiques et physiques par le biais de l'alimentation, ainsi que des expositions provenant d'autres sources, le cas échéant. (Manuel de procédure de la Commission du Codex Alimentarius) Aux fins du présent document, le terme «exposition alimentaire» désigne l'absorption d'une substance par une personne dans le cadre de son régime alimentaire (par le biais des aliments, des boissons, de l'eau potable et des compléments alimentaires).
Valeur extrême	Les variables les plus grandes et les plus petites parmi un échantillon d'observations. (Cambridge Dictionary of Statistics) Aux fins du présent document, les valeurs maximales et les valeurs proches d'un ensemble de données sont appelées valeurs extrêmes.
Valeur quantifiée	Un nombre qui n'est pas infini, c'est-à-dire qui peut être mesuré ou auquel on peut donner une valeur. Aux fins du présent document, on entend par valeur limitée tout résultat d'analyse égal ou supérieur à la LOQ déclarée.

Terme	Définition / Explication										
Groupe d'aliments	Ensemble de produits de base individuels présentant des caractéristiques biologiques et morphologiques similaires et donc un potentiel similaire de concentration d'une substance chimique préoccupante, et pour lesquelles une limite maximale commune peut être fixée.										
Distribution gamma	La distribution de probabilité, $f(x)$, donnée par $f(x) = \frac{x^{\gamma-1} \exp(-x\beta)}{\beta \Gamma(\gamma)}, \quad 0 \leq x < \infty, \beta > 0, \gamma > 0$ β est un paramètre d'échelle et γ un paramètre de forme. La moyenne, la variabilité, l'asymétrie et l'aplatissement de la distribution sont les suivants. $\text{mean} = \beta\gamma$ $\text{variance} = \beta^2\gamma$ $\text{skewness} = 2\gamma^{-\frac{1}{2}}$ $\text{kurtosis} = 3 + \frac{6}{\gamma}$ <table border="1"> <tr> <td>EN</td><td>FR</td></tr> <tr> <td>Mean</td><td>Moyenne</td></tr> <tr> <td>Variance</td><td>Variabilité</td></tr> <tr> <td>Skewness</td><td>Asymétrie</td></tr> <tr> <td>Kurtosis</td><td>Aplatissement</td></tr> </table> La distribution de $u = x/\beta$ est la distribution gamma standard avec la fonction de densité correspondante donnée par $f(u) = \frac{x^{\gamma-1} e^{-x}}{\Gamma(\gamma)}$ La fonction Γ définie par $\Gamma(r) = \int_0^\infty t^{r-1} e^{-t} dt$ Où $r > 0$ (r n'est pas nécessairement un nombre entier). La fonction est récursive et satisfait à la relation. $\Gamma(r+1) = r\Gamma(r)$ (Cambridge Dictionary of Statistics)	EN	FR	Mean	Moyenne	Variance	Variabilité	Skewness	Asymétrie	Kurtosis	Aplatissement
EN	FR										
Mean	Moyenne										
Variance	Variabilité										
Skewness	Asymétrie										
Kurtosis	Aplatissement										
GEMS/Aliments	Le Système mondial de surveillance de l'environnement – Programme de surveillance et d'évaluation de la contamination des aliments (Programme GEMS/Aliments) de l'Organisation mondiale de la santé – qui tient à jour des bases de données sur les niveaux de contaminants dans les aliments et les estimations de l'exposition alimentaire aux produits chimiques alimentaires. (OMS, EHC 240)										
Test de Grubbs	L'un des tests permettant d'identifier les valeurs aberrantes, qui est principalement un test de variabilité inter-laboratoires et peut également être utilisé (si $n > 2$) lorsque le test de Cochran a soulevé des doutes quant au fait que la forte variation intra-laboratoire est attribuable à un seul des résultats du test. (ISO 5725-2)										
Valeur indicative basée sur la santé (VABS)	Valeur numérique obtenue en divisant un point de départ (une dose sans effet indésirable observé, une dose de référence ou une limite de confiance inférieure de la dose de référence) par un facteur d'incertitude composite afin de déterminer un niveau qui peut être ingéré sur une période de temps définie (par exemple, toute une vie ou 24 heures) sans risque appréciable pour la santé. (OMS, EHC 240)										

Terme	Définition / Explication
Histogramme	Représentation graphique d'un ensemble d'observations dans laquelle les fréquences des classes sont représentées par les surfaces des rectangles centrés sur l'intervalle des classes. Si ces dernières sont toutes égales, les hauteurs des rectangles sont également proportionnelles aux fréquences observées. (Cambridge Dictionary of Statistics)
Apport journalier estimé international (IEDI)	Prévision de l'absorption journalière à long terme d'un résidu de pesticide sur la base des hypothèses de consommation alimentaire journalière moyenne par personne et des résidus médians provenant d'essais supervisés, en tenant compte des résidus dans la partie comestible d'une denrée et en incluant les composants de résidus définis par la Réunion conjointe FAO/OMS sur les résidus de pesticides pour l'estimation de l'apport alimentaire. Les modifications des niveaux de résidus résultant de la préparation, de la cuisson ou de la transformation commerciale sont incluses. Lorsque les informations sont disponibles, l'apport alimentaire de résidus provenant d'autres sources doit être inclus. L'IEDI est exprimé en milligrammes de résidus par personne. (OMS, EHC 240)
Estimation internationale de l'apport à court terme (IESTI)	Prévision de l'absorption à court terme d'un résidu de pesticide sur la base des hypothèses d'une consommation alimentaire quotidienne élevée par personne et des résidus les plus élevés provenant d'essais supervisés, en tenant compte des résidus dans la partie comestible d'un produit de base, et en incluant les composants de résidus définis par la Réunion conjointe FAO/OMS sur les résidus de pesticides pour l'estimation de l'apport alimentaire. L'IESTI est exprimé en milligrammes de résidus par kilogramme de poids corporel. (OMS, EHC 240)
Intervalle interquartile	Mesure de la dispersion donnée par la différence entre le premier quartile (25 ^e centile) et le troisième quartile (75 ^e centile) d'un échantillon. (Cambridge Dictionary of Statistics)
Comité mixte FAO/OMS d'experts des additifs alimentaires (JECFA)	Un comité d'experts qui se réunit depuis 1956. Le JECFA s'est engagé à collecter et à évaluer les données scientifiques sur les additifs alimentaires et à formuler des recommandations sur les niveaux d'utilisation sûrs. Cette tâche a été accomplie 1) en élaborant des spécifications relatives à l'identité et à la pureté des additifs alimentaires individuels qui ont été testés sur le plan toxicologique et qui sont commercialisés et 2) en évaluant les données toxicologiques relatives à ces additifs alimentaires et en estimant les doses acceptables pour l'homme. En 1972, le champ d'application des évaluations a été étendu aux contaminants dans les aliments, tandis qu'en 1987, le champ d'application a été encore élargi pour inclure les résidus de médicaments vétérinaires dans les aliments. (OMS, EHC 240)
Réunion conjointe FAO/OMS sur les résidus des pesticides (JMPR)	Titre abrégé de la réunion conjointe du groupe d'experts de la FAO sur les résidus de pesticides dans les aliments et l'environnement et du groupe principal d'évaluation de l'OMS sur les résidus de pesticides, qui se réunit depuis 1963. Les réunions sont normalement convoquées une fois par an. Le groupe d'experts de la FAO est chargé d'examiner les résidus et les aspects analytiques des pesticides considérés, y compris les données sur leur métabolisme, leur devenir dans l'environnement et leurs modes d'utilisation, et d'estimer les teneurs maximales en résidus et les teneurs médianes en résidus des essais contrôlés qui pourraient résulter de l'utilisation du pesticide conformément aux bonnes pratiques agricoles. Le groupe principal d'évaluation de l'OMS sur les résidus de pesticides est chargé d'examiner les données toxicologiques et connexes sur les pesticides et, dans la mesure du possible, d'estimer les apports alimentaires journaliers admissibles et les apports alimentaires à long terme de résidus. Le cas échéant, des doses de référence aiguës pour les pesticides sont estimées, ainsi que des estimations appropriées de l'apport alimentaire à court terme. (OMS, EHC 240)

Terme	Définition / Explication
Aplatissement	La mesure dans laquelle le pic d'une distribution de probabilité ou de fréquence unimodale s'écarte de la forme d'une distribution normale, en étant soit plus pointu (<i>leptokurtique</i>), soit plus plat (<i>platikurtique</i>). (Cambridge Dictionary of Statistics)
Test H de Kruskal-Wallis	Une méthode sans distribution qui est l'analogue de l'analyse de la variabilité d'un plan à sens unique. Elle permet de vérifier si les groupes à comparer ont la même médiane de population. La statistique de test est obtenue en classant toutes les N observations de 1 à N, quel que soit le groupe auquel elles appartiennent, puis en calculant $H = \frac{12 \sum_{i=1}^k n_i (\bar{R}_i - \bar{R})^2}{N(N-1)}$ Où n_i est le nombre d'observations dans le groupe i, $\bar{R}_i$ est la moyenne de leurs rangs, $\bar{R}$ est la moyenne de tous les rangs, donnée explicitement par $(N+1)/2$. Lorsque l'hypothèse nulle est vraie, la statistique du test a une distribution χ^2 avec $k-1$ degrés de liberté. (Cambridge Dictionary of Statistics)
Données censurées à gauche	Données qui ne sont pas des valeurs limitées (valeurs quantifiées) ou qui sont inférieures aux LOQ (ou LOD rapportée dans le cas où la LOQ n'a pas été rapportée).
Limite de détection (LOD)	La concentration ou la quantité nette réelle de l'analyte dans le matériau à analyser qui permettra, avec une probabilité $(1-\beta)$, de conclure que la concentration ou la quantité de l'analyte dans le matériau analysé est supérieure à celle du matériau vierge. Elle est définie comme suit: $\Pr(\hat{L} \leq LC L=LOD) = \beta$ où $\hat{L}$ est la valeur estimée, L est l'espérance ou la valeur réelle, et LC est la valeur critique. (CXG 72-2009) OU La limite de détection (LOD) est la concentration minimale d'un contaminant qui peut être mesurée qualitativement dans un aliment spécifique. La limite de détection est indiquée par un laboratoire ou une valeur calculée à partir de la LOQ. (Programme GEMS/Aliments)
Limite de quantification (LOQ)	Caractéristique de performance d'une méthode généralement exprimée en termes de signal ou de valeur de mesure (vraie) qui produira des estimations ayant un écart-type relatif (RSD) spécifié, généralement 10 % (ou 6 %). La LOQ est estimée par: $LOQ = kQ \sigma_Q$, $kQ = 1/RSDQ$ Où LOQ est la limite de quantification, σ_Q est l'écart-type à ce point, et kQ est le multiplicateur dont la réciproque est égale à la RSD choisie (La RSD approximative d'une estimation σ, basée sur v-degrés de liberté est $1/\sqrt{2v}$). (CXG 72-2009) OU La limite de quantification (LQ) est la concentration minimale d'un contaminant qui peut être mesurée quantitativement dans un aliment spécifique avec un niveau acceptable d'exactitude et de précision. La limite de quantification est indiquée par un laboratoire ou une valeur calculée à partir de la LOD. (Programme GEMS/Aliments)

Terme	Définition / Explication										
Distribution log-normale	La distribution de probabilité d'une variable aléatoire, X, pour laquelle $\ln(X)$ a une distribution normale avec une moyenne μ et une variabilité σ^2. La distribution est donnée par $f(x) = \frac{1}{x\sigma(2\pi)^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma^2} \{-(\ln x - \mu)^2\}\right] \quad 0 \leq x < \infty$ La moyenne, la variabilité, l'asymétrie et l'aplatissement de la distribution sont $\text{mean} = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$ $\text{variance} = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$ $\text{skewness} = (\exp(\sigma^2) + 2)(\exp(\sigma^2) - 1)^{\frac{3}{2}}$ $\text{kurtosis} = \exp(4\sigma^2) + 2\exp(3\sigma^2) + 3\exp(2\sigma^2) - 3$ <table border="1"> <tr> <td>EN</td><td>FR</td></tr> <tr> <td>Mean</td><td>Moyenne</td></tr> <tr> <td>Variance</td><td>Variabilité</td></tr> <tr> <td>Skewness</td><td>Asymétrie</td></tr> <tr> <td>Kurtosis</td><td>Aplatissement</td></tr> </table> Pour les petits, la distribution est approximée par la distribution normale. (Cambridge Dictionary of Statistics)	EN	FR	Mean	Moyenne	Variance	Variabilité	Skewness	Asymétrie	Kurtosis	Aplatissement
EN	FR										
Mean	Moyenne										
Variance	Variabilité										
Skewness	Asymétrie										
Kurtosis	Aplatissement										
Limite maximale (LM)	La limite maximale Codex pour un contaminant dans une denrée alimentaire ou un aliment pour animaux est la concentration maximale de cette substance recommandée par la Commission du Codex Alimentarius pour être légalement autorisée dans ce produit de base. (Manuel de procédure de la Commission du Codex Alimentarius)										
Test U de Mann-Whitney	Méthode sans distribution utilisée comme alternative au test T de Student pour évaluer si deux populations ont la même localisation. Étant donné un échantillon d'observations de chaque population, toutes les observations sont classées comme si elles provenaient d'un seul échantillon, et la statistique du test est la somme dans le plus petit groupe. Des tableaux donnant les valeurs critiques de la statistique du test sont disponibles et, pour des échantillons de taille moyenne ou grande, une approximation normale peut être utilisée. (Cambridge Dictionary of Statistics)										
Incertitude de mesure	Paramètre, associé au résultat d'un mesurage, qui caractérise la dispersion des valeurs qui pourraient raisonnablement être attribuées au mesurande (c'est-à-dire la grandeur à mesurer). (CXG54-2004, modifiée en 2021)										
Moyenne (Arithmétique)	Mesure de l'emplacement ou de la valeur centrale d'une valeur continue. Pour un échantillon d'observations $x_1, x_2, \dots, x_n$, la mesure est calculée comme $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ Très utile lorsque les données ont une distribution symétrique et ne contiennent pas de valeurs aberrantes. (Cambridge Dictionary of Statistics)										
Moyenne (Géométrique)	Une mesure de la localisation, g, calculée à partir d'un ensemble d'observations $x_1, x_2, \dots, x_n$ comme $g = \left(\prod_{j=1}^n x_j\right)^{\frac{1}{n}}$ (Cambridge Dictionary of Statistics)										

Terme	Définition / Explication
Médiane	La valeur d'un ensemble d'observations classées qui divise les données en deux parties de taille égale. Lorsque le nombre d'observations est impair, la médiane est la valeur intermédiaire. Lorsque le nombre d'observations est pair, la mesure est calculée comme la moyenne des deux valeurs centrales. Fournit une mesure de la localisation d'un échantillon qui convient aux distributions asymétriques et qui est également relativement insensible à la présence de valeurs aberrantes. (Cambridge Dictionary of Statistics)
Distribution multimodale	Une distribution de probabilité ou de fréquence qui a deux modes ou plus (pics). La multimodalité est souvent considérée comme une indication que la distribution observée résulte du mélange des distributions de groupes d'observations relativement distincts. (Cambridge Dictionary of Statistics)
Distribution normale	Une distribution de probabilité, $f(x)$ d'une variable aléatoire, X, qui est supposée par de nombreuses méthodes statistiques. Spécifiquement donné par $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right]$ Où μ et σ^2 sont, respectivement la moyenne et la variabilité de x. (Cambridge Dictionary of Statistics)
Valeur aberrante	Un membre d'un ensemble de valeurs qui est incompatible avec d'autres membres de cet ensemble. (CXG72-2009)
Test paramétrique/test non paramétrique	Les tests paramétriques supposent que les données suivent un certain modèle de distribution, qui est généralement une distribution normale. Les tests non paramétriques ne supposent pas que les données aient une distribution particulière et peuvent analyser des données pour lesquelles aucun test paramétrique n'est applicable.
Centile	L'ensemble des divisions qui produisent exactement 100 parties égales dans une série de valeurs continues, telles que la concentration d'un certain contaminant dans les aliments. Par exemple, un échantillon dont la concentration est supérieure au 95 ^e centile présente une concentration plus élevée que plus de 95 % des autres niveaux de contaminants. (Cambridge Dictionary of Statistics, modifié pour les besoins de ces directives)
Pays producteur primaire	Un pays où les produits de base sont cultivés et récoltés
Dose journalière maximale tolérable provisoire (DJMP) / Dose journalière tolérable provisoire (DJTP)	Un type de VABS. Le seuil de risque utilisé pour les contaminants dont les propriétés ne sont pas cumulatives. Sa valeur représente l'exposition humaine admise suite à la contamination naturelle des aliments et de l'eau potable par la substance. Pour les micro-constituants qui sont à la fois des nutriments essentiels et des constituants involontaires, une fourchette est donnée, dont la valeur inférieure représente le niveau d'essentialité et la valeur supérieure la DJMTP. (CXS193-1995)
Dose hebdomadaire tolérable provisoire (DHTP)	Un type de VABS. Le seuil utilisé pour les contaminants alimentaires comme les métaux lourds dont les propriétés sont cumulatives. Sa valeur représente l'exposition à ces contaminants hebdomadaire admise pour l'homme, involontairement associée à la consommation d'aliments autrement sains et nutritifs. (CXS193-1995)
Dose mensuelle tolérable provisoire (DMTP)	Un type de VABS. Le seuil utilisé pour un contaminant alimentaire dont les propriétés cumulatives ont une demi-vie longue durée dans le corps humain. Sa valeur représente l'exposition à un contaminant mensuelle admise pour l'homme, involontairement associé à des aliments autrement sains et nutritifs. (CXS193-1995)

Terme	Définition / Explication
Assurance qualité (dans les laboratoires d'analyse)	Toutes les actions planifiées et systématiques nécessaires pour donner une confiance adéquate dans le fait que les résultats analytiques satisferont aux exigences de qualité données (CXG72-2009) Il existe un certain nombre de recommandations du Codex sur l'assurance qualité basées sur les considérations du Comité du Codex sur les méthodes d'analyse et d'échantillonnage (par exemple, CXG27-1997, CXG28-1995, CXG64-1995 et CXG65-1997)
Taux de rejet/taux d'infraction	Taux de rejet (ou taux d'infraction) (%) = (nombre de points de données présentant des concentrations d'un contaminant supérieures à la LM)/nombre total de points de données x 100.
Analyse de sensibilité	L'analyse de sensibilité montre comment différentes valeurs d'une variable indépendante affectent une variable dépendante dans le cadre d'un ensemble donné d'hypothèses. Dans le cas des données d'occurrence, il est utilisé pour évaluer l'impact de l'exclusion de données sur, par exemple, les valeurs moyennes, médianes et du centile élevées.
Asymétrie	L'absence de symétrie dans une distribution probabiliste. (Cambridge Dictionary of Statistics)
Écart-type	La mesure la plus couramment utilisée de la dispersion d'un ensemble d'observations. Égale à la racine carrée de la variabilité s^2 , qui est donnée par la formule suivante: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ où $x_1, x_2, \dots, x_n$ sont les observations de l'échantillon n et $\bar{x}$ est la moyenne de l'échantillon. (Cambridge Dictionary of Statistics)
Puissance statistique	En statistique, la puissance d'un test d'hypothèse binaire est la probabilité que le test rejette correctement l'hypothèse nulle lorsqu'une hypothèse alternative spécifique est vraie. L'hypothèse nulle du test U de Mann-Whitney ou du test H de Kruskal-Wallis est que chaque ensemble de données provient de la même population.
Échantillonnage aléatoire	Est un type d'échantillonnage. Le terme d'«échantillonnage aléatoire» devrait être choisi pour l'échantillonnage de routine, même s'il cible des types d'aliments spécifiques ou des pays importateurs spécifiques. Les tests effectués sur une large gamme d'échantillons importés d'une certaine catégorie d'aliments pour détecter la présence d'un contaminant donné seraient «aléatoires».
Échantillonnage ciblé	Est un type d'échantillonnage. Le terme d'«échantillonnage ciblé» devrait être choisi pour les échantillonnages de suivi réalisés à la suite de constatations spécifiques de contamination. Par exemple, si un pays identifie des niveaux élevés d'un contaminant dans un échantillon provenant d'un fabricant particulier, un échantillonnage supplémentaire du même lot ou des lots produits simultanément par le même fabricant serait «ciblé».

APPENDICE II

Questions identifiées en vue d'une éventuelle discussion future

Nombre minimum d'échantillons pour l'estimation des valeurs de centile élevées avec un niveau de confiance élevé (REP23/CF16, par. 93 et 94, par. 98 (vi) (a)).

- Une orientation supplémentaire reste à fournir en ce qui concerne l'ensemble de données sur lequel la LM devrait être basée ou sur la base de données à laquelle il conviendrait d'accorder la priorité pour l'élaboration de LM (ensemble de données combiné, ensemble de données indiquant les schémas de contamination les plus élevés pour autant que le produit de base ait été produit selon de bonnes pratiques, ensembles de données émanant des principaux pays ou régions de production, ensembles de données émanant de pays importateurs reflétant les taux d'un contaminant présent dans un produit de base faisant l'objet d'un commerce international, ensemble de données à utiliser à décider au cas par cas) (REP23/CF16, par. 98 (vi) (b)).
- Examiner plus avant le rôle du Comité concernant le calcul des taux de réduction de l'exposition alimentaire lors de l'examen des LM. (Le calcul de l'exposition alimentaire est une fonction d'évaluation des risques qui devrait être entreprise par le JECFA, et le JECFA fournit l'avis scientifique sur lequel les décisions de gestion des risques prises par le Comité sont basées - il est important de clarifier les rôles du JECFA et du CCCF en tant que, respectivement, évaluateur et gestionnaire des risques, dans le calcul des taux de réduction de l'exposition alimentaire lors de l'examen des LM (REP23/CF16, par. 90 et 91, et par. 98 (vi) (c)). (voir aussi le paragraphe 17 (ii) des Recommandations au CCCF18 – au cas où une conclusion serait atteinte sur ce sujet lors du CCCF18, aucuns travaux futurs ne seraient nécessaires).
- Processus plus structuré pour l'élaboration des appels de données; (REP23/CF16, par. 98 (viii)).
- Identification des taux de rejet appropriés lors de l'établissement de LM (orientation relative aux éléments à prendre en compte pour définir le taux de rejet approprié) (CX/CF 22/15/14, chapitre IV de l'Appendice I, CF16/CRD06, par. 32).
- Pertinence des régimes alimentaires par module de consommation basés sur le marché de GEMS/Aliments pour l'élaboration des LM (concilier les estimations réalistes provenant des données nationales sur la consommation avec les données relatives à la disponibilité/l'utilisation/le marché dans les régimes alimentaires par module de GEMS/Aliments (p. ex. sucre, épices/herbes, thés, cafés).
- Élaboration d'Orientation sur la pondération des données dans les analyses où le nombre de points de données est significativement différent entre les ensembles de données individuels provenant de différents pays ou régions.

APPENDICE III**Liste des participants****PRÉSIDENCE****Union européenne**

Frans Verstraete
European Commission

COPRÉSIDENTS**Japon**

Tetsuo Urushiyama
Ministry of Agriculture, Forestry and Fisheries

Yoshiyuki Takagishi
Ministry of Agriculture, Forestry and Fisheries

Pays-Bas

Nikki Emmerik
Ministry of Health, Welfare and Sport

Moniek Ringenier
Ministry of Health, Welfare and Sport

États-Unis d'Amérique

Lauren Posnick Robin
U.S. Food and Drug Administration

Eileen Abt
U.S. Food and Drug Administration

PAYS MEMBRES ET ORGANISATIONS MEMBRES**ARGENTINE**

Maria Alejandra Rodriguez
Instituto Nacional de Tecnologia Industrial

Martin Fernández
Instituto Nacional de Alimentos

Gisele Simondi
Instituto Nacional de Alimentos

Silvana Ruarte
Instituto Nacional de Alimentos

AUSTRALIE

Matthew Joseph O'Mullane
Food Standards Australia New Zealand

Nick Fletcher
Food Standards Australia New Zealand

BELGIQUE

Christine Vinkx
FPS Health, Food Chain Safety and Environment

Andrea Carletta
FPS Health, Food Chain Safety and Environment

Fabio Enrico Occhetti
Federal Agency for the Safety of the Food Chain

BRÉSIL

Larissa Bertollo Gomes Porto
ANVISA

Giselle Kindlein
Ministry of Agriculture

BRUNEI

Nur Halimatussaadiah Mohd Alias
Brunei Darussalam Food Authority

CANADA

Ian Richard
Health Canada | Santé CANADA

Carla Hilts
Health Canada

Stephanie Glanville
Health Canada

Rosalie Awad
Health Canada

CHILI

Lorena Delgado Rivera
Instituto de Salud Pública de Chili

CHINE

Yi Shao
China National Center of Food Safety Risk Assessment (CFSA)

Yongning Wu
National Center of Food Safety Risk Assessment (CFSA)

Dawei CHEN

Shuang Zhou

National Center of Food Safety Risk Assessment (CFSA)

ÉQUATEUR

Saul Flores

AGROCALIDAD

UNION EUROPÉENNE

Veerle Vanheusden

European Commission

FRANCE

Jean-Cédric Reninger

ANSES.

Niels Enslen

Ministry of Agriculture

INDE

Reeba Abraham

Ministry of Commerce and Industry (APEDA)

Aditya Jain

National Dairy Development Board

INDONÉSIE

Yeni Restiani

Indonesian Food and Drug Authority

IRAN

Mansoor Mazaheri

ISIRI-Standard Research Institute

JAPON

Daisuke Aoki

Consumer Affairs Agency

Mio Toda

National Institute of Health Sciences

MEXIQUE

Tania Daniela Fosado Soriano

Secretaría de Economía

Codex Committee on Fresh Fruits and Vegetables (CCFFV)
secretariat

CCFFV

MAROC

Ouazzani Sanae

ONSSA

PAYS-BAS

Zihui Chen

RIVM

Astrid Bulder

RIVM

NOUVELLE-ZÉLANDE

Jeane Nicolas

Ministry for Primary Industries, New Zealand Food

Fiapaipai Ruth Auapaau

Ministry for Primary Industries

NORVÈGE

Olga Soleng

Norwegian Food Safety Authority

NIGÉRIA

Abu Rachel Kakataidii

NAFDAC

Mazai Maymunah Ummjamil

Standards Organisation of Nigeria

PHILIPPINES

Phelan Apostol

Food and Drug Administration Philippines

RÉPUBLIQUE DE CORÉE

Republic of Korea/ Codex Secretariat

Ministry of Agriculture, Food and Rural Affairs

Kim Jooyeon

Ministry of Food and Drug Safety

Heewon Moon

Ministry of Food and Drug Safety

FÉDÉRATION DE RUSSIE

Darya Bagreeva

Federal Scientific Center of Hygiene

RWANDA

Rosine Niyonshuti

Rwanda Standards Board

ARABIE SAOUDITE

Mohammed Mousa Ali Alshehri

Saudi Food and Drug Authority

Lama Almainan

Saudi Food and Drug Authority

SÉNÉGAL

Mame Diarra Faye

Ministère de la Santé et l'Action sociale

SINGAPOUR

How Chee Ong

Singapore Food Agency

Er Jun Cheng

Singapore Food Agency

Wesley Zongrong Yu

Singapore Food Agency

AFRIQUE DU SUD

Juliet Masuku

Department of Health

SOUDAN

Liza Nelson Michael Taban

Sudan Bureau of Standards

SUÈDE

Nurun Nahar

Swedish Food Agency

SUISSE

Judit Valentini
Federal Food Safety and Veterinary Office

THAÏLANDE

Chutiwan Jatupornpong
Ministry of Agriculture and Cooperatives, National Bureau of
Agricultural Commodity and Food Standards

TURQUIE

Bengi Akbulut Pınar
Ministry of Agriculture and Forestry

Sinan Arslan
Ministry of Agriculture and Forestry

UGANDA

Francis Enaru
Ministry of Trade, Industry and Cooperatives

ROYAUME-UNI

Mark Willis
Food Standards Agency

Craig Jones
Food Standards Agency

Helen Twyble
Food Standards Agency

Izaak Fryer-Kanssen
Food Standards Agency

Holly Howell-Jones
Food Standards Agency

Clare McCartney-Collard
Food Standards Agency

ÉTATS-UNIS D'AMÉRIQUE

Sara McGrath
U.S. Food and Drug Administration

Linda Benjamin
U.S. Food and Drug Administration

Quynh-Anh Nguyen
FDA/CFSAN

Alexandra Ferraro
U.S. Codex Office | U.S. Department of Agriculture

Armen Mirzoian
Alcohol and Tobacco Tax and Trade Bureau (TTB) – Scientific
Services Division

URUGUAY

Santiago Viera
Laboratorio Tecnológico del Uruguay

Natalie Merlinski
Ministry of Livestock, Agriculture and Fisheries

ZAMBIE

Henry Njapau
Zambia Academy of Sciences

NON-GOVERNMENTAL ORGANIZATIONS**INTERNATIONAL COUNCIL OF BEVERAGES ASSOCIATIONS**

Simone SooHoo

Maia Jack

INTERNATIONAL ORGANISATION OF SPICE TRADE ASSOCIATIONS

Shannen Kelly

INTERNATIONAL SPECIAL DIETARY FOODS INDUSTRIES

Marian Brestovansky

Jean Christophe Kremer

FOODDRINKEUROPE

Alejandro Rodarte

THIE | TEA & HERBAL INFUSIONS EUROPE

Farshad Rostami

FEDIOL | EU VEGETABLE OIL AND PROTEINMEAL INDUSTRY ASSOCIATION

Kalila Hajjar

AMERICAN PEANUT COUNCIL

Jim Elder