



PROGRAMME MIXTE FAO/OMS SUR LES NORMES ALIMENTAIRES

COMITÉ DU CODEX SUR LES CONTAMINANTS DANS LES ALIMENTS

Dix-septième session
15-19 avril 2024
Panama (ville), Panama

LIGNES DIRECTRICES POUR L'ANALYSE DES DONNÉES EN VUE DE L'ÉLABORATION DE LIMITES MAXIMALES ET DE L'AMÉLIORATION DE LA COLLECTE DES DONNÉES (*Document préparé par le groupe de travail électronique présidé par l'Union européenne et coprésidé par le Japon, les Pays-Bas et les États-Unis d'Amérique*)

CONTEXTE

1. Lors de la 12^e session¹, le CCCF a examiné la proposition du Secrétariat du JECFA de rédiger des lignes directrices générales pour l'analyse des données en vue de l'élaboration de LM, car il a été remarqué que les GTE adoptaient différentes approches. Ces différences se rapportaient, par exemple, au traitement des données d'occurrence sans information sur la LOQ. Des lignes directrices générales aideraient les futurs GTE à adopter des approches cohérentes en matière d'analyse des données. Le CCCF est 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 lignes directrices générales pour l'analyse des données en vue de l'établissement de LM a été présenté et le CCCF est 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 du CCCF (CCCF14), le CCCF est 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 du document CX/CF 21/14/15, a été diffusée en octobre 2021, assortie d'une demande de commentaires sur les lignes directrices pour l'analyse des données en vue de l'établissement de limites maximales et pour l'amélioration de la collecte des données.
4. Le document CX/CF 21/14/15 a été mis à jour pour prendre en compte les commentaires reçus en réponse à la lettre circulaire CL 2021/78 CF, ainsi que ceux mentionnés lors de la 14^e session du CCCF. En conséquence, le document a fait l'objet d'une révision significative, mettant également en lumière la nécessité de le restructurer. Étant donné la disponibilité tardive du document et en tenant compte des commentaires reçus et des changements significatifs proposés, il restait trop peu de temps pour une discussion des coprésidents et pour qu'ils puissent apporter leur contribution sur un document à faire circuler pour commentaires. Le document mis à jour a été joint en appendice I au document CX/CF 22/15/14 uniquement pour information.
5. Une réunion parallèle virtuelle s'était tenue préalablement à la 15^e session⁵ du CCCF (CCCF15) afin de discuter du plan de travail pour l'année prochaine et de certains aspects du contenu des lignes directrices, en particulier la structure et les sujets devant leur être inclus.
6. La 15^e session du CCCF est convenue:

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

- a) de la convocation de trois réunions virtuelles des groupes de travail en 2022 en vue d'obtenir des contributions et de développer le document;
 - b) de la création de trois sous-groupes présidés par les coprésidents, et de la division suivante des sujets devant être discutés au sein des trois sous-groupes:
 - tous les sujets liés à la collecte et à la soumission des données et à l'extraction des données de la base de données GEMS/Aliments,
 - tous les sujets liés à la sélection de données/au nettoyage des données et à la génération d'aperçus de données (aspect de l'analyse des données),
 - tous les sujets liés à l'analyse statistique (aspect de l'analyse des données), et
 - les aspects relatifs à la présentation des données sont étroitement liés à l'analyse des données et doivent par conséquent être examinés en connexion avec l'analyse des données dans les sous-groupes pertinents.
 - c) que le contenu des trois réunions virtuelles des groupes de travail reflèterait la division des sujets entre les trois sous-groupes;
 - d) du statut, des buts/objectifs et des utilisateurs cibles devant être présentés dans le préambule du document d'orientation;
 - e) 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. Le document de base à utiliser lors des réunions virtuelles des groupes de travail et des sous-groupes serait constitué du document figurant à l'appendice I du document CX/CF 22/15/14 divisé en trois parties séparées, suivant les responsabilités des sous-groupes, pour une discussion lors des réunions virtuelles des groupes de travail /sous-groupes
 - f) de rétablir le GTE présidé par l'UE et coprésidé par le Japon, les Pays-Bas et les États-Unis d'Amérique, travaillant en anglais, étant entendu que trois sous-groupes seront créés au sein du GTE, afin d'élaborer une proposition de lignes directrices générales pour l'analyse des données en vue de l'établissement de LM et de l'amélioration de la collecte de données.
7. L'appendice I du document CX/CF 22/15/14 «Proposition de lignes directrices pour l'analyse des données en vue de l'établissement de limites maximales et de l'amélioration de la collecte des données» a été communiquée au GTE afin qu'il fasse part de ses observations avant le 1^{er} octobre 2022.
 8. La première réunion virtuelle de groupe de travail s'est tenue le 11 octobre 2022 sous la présidence des Pays-Bas et a porté sur la sélection/le nettoyage des données, ainsi que sur la génération d'aperçus de données. La deuxième réunion virtuelle de groupe de travail, présidé par les États-Unis d'Amérique, s'est tenue le 19 octobre 2022 et a porté sur la collecte et la soumission des données, ainsi que sur leur extraction. La troisième réunion virtuelle de groupe de travail, présidé par le Japon, s'est tenue le 20 octobre 2022 et a porté sur l'analyse statistique.
 9. L'avant-projet de lignes directrices divisé en trois parties distinctes (c'est-à-dire a) collecte et soumission des données, et extraction des données de la base de données GEMS/Aliments, b) sélection/nettoyage des données et génération d'un aperçu de données (aspect de l'analyse des données) et c) analyse statistique (aspect de l'analyse des données), mis à jour par les présidents respectifs des groupes de travail virtuels, en tenant compte des discussions qui ont eu lieu lors des réunions virtuelles des groupes de travail et des commentaires reçus, a été diffusé au GTE pour commentaires.
 10. L'avant-projet de lignes directrices figurant à l'appendice IV du document CX/CF 23/16/12 est le résultat de la compilation, réalisée par le Président du GTE, des trois parties mises à jour en un seul et même document. En annexe à l'avant-projet de lignes directrices, un glossaire est fourni. Le Président du GTE ayant mis le document très tardivement à disposition, celui-ci n'a pas été diffusé pour commentaires et a uniquement été fourni pour information à la 16^e session du CCCF.
 11. Lors de la 16^e session du CCCF, un accord a été trouvé concernant les changements à apporter à la base de données GEMS/Aliments, présentés sous forme de recommandations à l'administrateur du GEMS/Aliments pour révision (appendice I)⁶. Une fois que ce dernier aura indiqué les recommandations susceptibles d'être mises en œuvre de manière efficace et le calendrier de cette mise en œuvre, la rubrique «collecte, soumission et extraction des données» devra être mise à jour en tenant compte de l'avis de l'administrateur de la base de

⁶ REP23/CF16 par. 98 (i) and (ii)

données GEMS/Aliments. La rubrique mise à jour sera communiquée au GTE pour commentaires et finalisée pour soumission au Secrétariat du Codex, en vue de sa diffusion pour commentaires et de sa finalisation lors de la 17^e session du CCCF.

12. Lors de la 16^e session du CCCF, il a été convenu de mettre à jour les rubriques «extraction/sélection/nettoyage des données» et «analyse des données», contenant les éléments et principes de base, et de communiquer les rubriques mises à jour au GTE pour commentaires, puis de soumettre le résultat de la consultation du GTE au Secrétariat du Codex en vue de sa diffusion pour commentaires, dans la perspective d'un éventuel accord provisoire lors de la 17^e session du CCCF.⁷ En outre, il a été convenu qu'une liste des thèmes des rubriques «sélection/nettoyage des données – génération d'un aperçu de données» et «analyse statistique» sera élaborée pour examen et accord par la 17^e session du CCCF, en vue d'une discussion ultérieure après cette 17^e session⁸.

⁷ REP23/CF16 par. 98 (iii) et (iv)

⁸ REP23/CF16, par. 98 (v)

TRAVAUX NON RÉALISÉS / RÉALISÉS DEPUIS LA 16^e SESSION DU CCCF

13. Toutefois, l'administrateur de GEMS/Aliments a été consulté beaucoup trop tard par le Président du GTE et, par conséquent, la rubrique «collecte et soumission des données» n'a pas pu être mise à jour en tenant compte de ses observations ni communiquée au GTE pour commentaires ni soumise à la 17^e session du CCCF en vue de sa finalisation. La rubrique «collecte/soumission/extraction des données» des lignes directrices figure, dans sa version actuelle, à l'appendice II. Cette rubrique devra être mise à jour une fois que l'administrateur de GEMS/Aliments aura donné son avis sur la faisabilité, et que la 16^e session du CCCF aura accepté les changements recommandés.
14. Les rubriques «extraction/sélection/nettoyage des données» et «analyse des données» n'ont pas été mises à jour par le Président du GTE et n'ont pas été soumises au GTE pour commentaires. Par conséquent, aucun document actualisé des rubriques «sélection/nettoyage des données - génération d'aperçus de données» et «analyse statistique» n'est soumis à la 17^e session du CCCF en vue d'une discussion et d'un éventuel accord provisoire.
15. Une liste des titres/sujets à aborder dans les rubriques «extraction/sélection/nettoyage des données» et «analyse des données» pour une éventuelle discussion lors de la 17^e session du CCCF est fournie à l'appendice III du présent document.
16. L'appendice IV du présent document contient, à titre d'information uniquement, les rubriques «extraction/sélection/nettoyage des données» et «analyse des données» telles que présentées à l'appendice IV du document CX/CF 23/16/12, dans lesquelles les résultats des discussions de la 16^e session du CCCF ont été intégrés.

POINTS DE DISCUSSION

17. Discussion sur les changements apportés à la base de données GEMS/Aliments, présentés sous forme de recommandations à l'administrateur de GEMS/Aliments pour révision, comme indiqué à l'appendice I, sur la base des réactions préliminaires de l'administrateur de la base de données GEMS/Aliments. En raison de la consultation très tardive de l'administrateur de la base de données GEMS/Aliments par le Président du GTE, le retour d'information final sera fourni après la 17^e session du CCCF.
18. Discussion relative aux sujets des rubriques «sélection/nettoyage des données» et «analyse des données» tels que figurant à l'appendice III, notamment la proposition de fusion/combinaison de certaines parties de la rubrique «analyse des données» avec des sujets de la rubrique «sélection/nettoyage des données».
19. Discussion sur le niveau de détail/la complexité de certaines parties de la rubrique «analyse des données» nécessaires dans le cadre des lignes directrices.
20. Discussion sur les sujets identifiés et répertoriés à l'appendice III, pour lesquels une discussion plus approfondie sera nécessaire à l'avenir (c'est-à-dire après la 17^e session du CCCF).
21. Confirmation de l'intégration correcte des résultats des discussions de la 16^e session du CCCF dans les rubriques «sélection/nettoyage des données» et «analyse des données».

CONCLUSION

22. Étant donné que les travaux sur l'avant-projet de lignes directrices pour l'analyse des données en vue de l'établissement de limites maximales (LM) et de l'amélioration de la collecte de données n'ont pas progressé comme prévu à l'origine, en raison de l'inactivité du Président du GTE, il est proposé de discuter d'une autre méthode de travail lors de la 17^e session du CCCF.

RECOMMANDATIONS

23. Il convient de noter que les recommandations ne portent pas sur le contenu des lignes directrices, mais sur l'approche à adopter à partir de maintenant pour permettre au CCCF de progresser dans ses travaux.
24. Étant donné que les lignes directrices pour l'analyse des données sont élaborées aux fins du processus de travail interne du CCCF, il est proposé de confier ces travaux non pas à un GTE, mais à un groupe de travail présession (qui pourrait fonctionner en mode présentiel ou virtuel) ou à un groupe de travail intersession (dénommé ci-après «le GT»), à l'instar des GT chargés des travaux du JECFA et de la révision des normes. Au sein de ce GT, les lignes directrices pourront être développées plus avant, et d'autres questions liées à l'analyse des données, émergeant de discussions sur des points de l'ordre du jour pourront également être abordées, le cas échéant. À cet effet, l'avant-projet de lignes directrices convenu sera inclus dans le rapport de la réunion du CCCF.

25. Une lettre circulaire pourrait être publiée pour solliciter des observations sur l'avant-projet de lignes directrices. Les commentaires seront compilés dans un document de travail et inclus dans l'ordre du jour. Le Président du GT préparera une proposition pour discussion au sein du GT, après quoi des recommandations pour la révision de l'avant-projet de lignes directrices seront présentées à la session plénière du CCCF. Un nouvel avant-projet approuvé sera ensuite inclus dans le rapport de la réunion. Un Président sera désigné lors de la 17^e session du CCCF.

LISTE DES APPENDICES

- APPENDICE I: Modifications de la base de données GEMS/Aliments recommandées par la 16^e session du CCCF, soumises à l'administrateur de GEMS/Aliments pour qu'il détermine lesquelles des recommandations peuvent être mises en œuvre de manière efficace, ainsi que le calendrier de mise en œuvre.
- APPENDICE II: La rubrique «collecte, soumission et extraction des données» des lignes directrices, dans sa version actuelle (**uniquement à titre d'information**)
- APPENDICE III: Une liste des titres/sujets à aborder/inclure dans les rubriques «sélection/nettoyage des données» et «analyse des données»
- APPENDICE IV. Les rubriques «sélection/nettoyage des données» et «analyse des données» dans lesquelles les résultats des discussions de la 16^e session du CCCF ont été intégrés, assortis d'un glossaire des termes utilisés (**uniquement à titre d'information**)
- APPENDICE V. Liste des membres du groupe de travail électronique (GTE)

APPENDICE I

Modifications proposées de la base de données GEMS/Aliments (soumises à l'administrateur de la base de données GEMS/Aliments pour examen/acceptation)

(pour examen)

Partie A: Modifications des champs existants – les champs sur fond gris sont ceux pour lesquels aucune modification n'est proposée.

Col	Champ	Type de champ/ Options du menu déroulant	Obligatoire ou Facultatif	Bloc de texte	Nouveau texte requis	Justification
E	Identifiant local de l'aliment	Texte libre	Obligatoire		Ajouter l'indicateur <i>sur la Fiche 2: Mappage de l'aliment</i> : Fournissez un nom détaillé dans l'Identifiant local de l'aliment, par exemple «Hoplostète orange» au lieu de «Poisson».	Note: Ceci vise à inciter les utilisateurs à saisir des noms qui seront plus utiles lors du tri et de l'analyse.
F	N° de série de l'enregistrement	Texte libre	Obligatoire		Ajouter le bloc de texte: «Un numéro de série (identifiant de l'échantillon) est utilisé pour chaque échantillon. Les données relatives à différents contaminants présents dans le même échantillon doivent porter le même numéro de série.»	Clarifie le n° de série de l'enregistrement.
G	Pays/Région	Menu	Facultatif		Modifier le nom du champ en «Soumettre pays/région» et/ou Ajouter le bloc de texte: «Correspond aux pays ou régions qui soumettent des données; il ne s'agit pas du pays de production.»	Est plus clair pour les personnes soumettant des données.
H	Contaminant	Menu	Facultatif	Actuel bloc de texte: «Veuillez sélectionner un contaminant dans la liste. Ceci est facultatif si un contaminant est indiqué à la première page.»	Bloc de texte modifié: «Veuillez sélectionner un contaminant dans la liste. Un contaminant est requis, mais saisie manuelle dans la colonne H: Le contaminant est facultatif si un contaminant a été indiqué dans la fiche 1: Début.»	Il est demandé de clarifier le texte du bloc, car des questions ont été posées sur la raison pour laquelle un contaminant est facultatif.
I	Origine de l'aliment	Menu: • Production	Facultatif			

		intérieure <ul style="list-style-type: none"> • Importation • Origine mixte • Pas d'informations 				
J	Date d'échantillonnage	Texte libre (JJJJ)	Obligatoire			
K	Représentativité/fiabilité de l'échantillon	Menu <ul style="list-style-type: none"> • Échantillonnage aléatoire • Échantillonnage ciblé • Pas d'informations 	Obligatoire		<p><i>Modifier le titre du champ:</i> Représentativité de l'échantillon</p> <p><i>Modifier le menu déroulant:</i> – pas de changement du menu déroulant</p> <ul style="list-style-type: none"> • Échantillonnage aléatoire • Échantillonnage ciblé • Pas d'informations <p>Ajouter un bloc de texte pour clarifier les termes «échantillonnage aléatoire» et «échantillonnage ciblé», et donner des explications dans les instructions pour la soumission électronique – se référer aux définitions des termes du glossaire.</p>	<p>Note: Il est demandé de supprimer «fiabilité» du nom du champ et d'ajouter (routine) après «aléatoire» dans le champ du menu déroulant.</p> <p><i>(Clarification proposée: 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».</i></p> <p><i>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 taux 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»).</i></p>
L	Identification en laboratoire	Texte libre	Facultatif			
M	Assurance qualité analytique	Menu <ul style="list-style-type: none"> • Assurance qualité interne uniquement • Tests d'aptitude réussis • Officiellement accrédité 	Facultatif			
N	Unités de mesure	Menu déroulant	Obligatoire		<ul style="list-style-type: none"> • mg/kg 	Ce champ est déjà obligatoire et les

	pour les taux de contaminants	<ul style="list-style-type: none"> • mg • ug • ng • pg • bg 			<ul style="list-style-type: none"> • (µg/kg) • ng/kg • pg/kg • Bq/kg 	unités complètes sont actuellement indiquées dans l'indicateur. Il est demandé que les unités complètes (mg/kg vs mg) apparaissent également dans les lignes.
O	LOD	Texte libre	Obligatoire pour les résultats non quantifiés si la LOQ n'est pas fournie		Facultatif Modifier l'ordre des champs: le champ O doit venir après le champ P	Note: Ceci ne peut devenir facultatif que si la LOQ est obligatoire.
P	LOQ	Texte libre	Obligatoire pour les résultats non quantifiés si la LOD n'est pas fournie		Obligatoire Modifier l'ordre des champs: le champ P doit venir après le champ O.	«Obligatoire» remplacerait «Obligatoire pour les résultats non quantifiés si la LOD n'est pas fournie ». (La mention «obligatoire» ne s'appliquerait qu'aux données nouvellement saisies afin de maintenir la validité des données précédemment soumises, sans déclaration de la LOQ).
Q	Résultats basés sur	Menu déroulant <ul style="list-style-type: none"> •Teneur en graisses •Poids sec •Tel quel (cru, frais) •Tel que consommé 	Obligatoire		<i>Modifier le menu déroulant comme suit:</i> <ul style="list-style-type: none"> • Tel quel (cru, frais, tel que vendu) • Tel que consommé • Teneur en graisses <ul style="list-style-type: none"> ---- % de graisses [texte libre, permettre % spécifique # ou fourchette] envisager d'indiquer cette information (%) dans un nouveau champ «informations sur la composition» • Poids sec 	Note: Il est demandé d'apporter des modifications au menu déroulant.

					---- % d'eau [texte libre, permettre % spécifique # ou fourchette] envisager d'indiquer cette information (%) dans un nouveau champ «informations sur la composition»	
R	Portion analysée	Menu <ul style="list-style-type: none"> •Uniquement comestible •Aliment total (comestible + non comestible) 	Obligatoire		<p><i>Modifier le menu déroulant comme suit:</i></p> <ul style="list-style-type: none"> •Uniquement comestible •Aliment entier (comestible + non comestible) <p><i>Ajouter à l'indicateur:</i></p> <p>Exemple: noix décortiquée (comestible) contre noix non décortiquée (aliment entier)</p>	Ce champ existe déjà et est déjà obligatoire. Il est demandé d'ajouter des exemples dans l'indicateur tels que «décortiqué versus non décortiqué/pelé versus non pelé» et de remplacer «total» par «entier».
S	État de l'aliment analysé	Menu <ul style="list-style-type: none"> •Cuit •Cru •Pas d'informations 	Facultatif		<p><i>Modifier le titre du champ comme suit:</i></p> <p>État de l'aliment analysé (Cuit/Cru)</p>	Il est demandé de préciser si ce champ s'applique, par exemple, à du poisson cuit ou à du poisson cru.
T	Résultats	Texte libre	Obligatoire	Indicateur actuel: Le résultat est obligatoire si la LOD et la LOQ ne sont pas fournies.	<p><i>Modifier le titre du champ comme suit:</i></p> <p>«Le résultat numérique est obligatoire si la LOD et la LOQ ne sont pas fournies.»</p>	Pour plus de clarté. Concerne la proposition de modification du champ O en «obligatoire». Bien qu'il ne soit plus pertinent pour les données nouvellement soumises si la modification du champ O est acceptée, l'indicateur reste pertinent pour les ensembles de données déjà soumis.
U	Échantillon global	Menu <ul style="list-style-type: none"> •Individuel 	Facultatif		Il est proposé de rendre ce champ obligatoire.	

		•Global				
V	Confidentialité des données	Menu •Oui •Vide	Facultatif		<i>Modifier le menu déroulant comme suit:</i> •Oui •Non	Pour plus de clarté; la signification de «vide» n'est pas claire.
W	Remarques/ Références	Texte libre	Facultatif			

Partie B: Nouveaux champs proposés:

Col	Champ proposé:	Type de champ/ Options du menu déroulant	Obligatoire ou Facultatif	Bloc de texte	Nouveau texte requis	Justification
--	Année de production	Texte libre (JJJJ)	Facultatif		S/O – nouveau champ	Facultatif – peut ne pas être connu
--	Informations sur la composition	Texte libre	Facultatif	Informations figurant sur les étiquettes, telles que les principaux ingrédients ou le pourcentage total de solides de cacao dans le chocolat Voir champ Q: ajouter la teneur en graisses ou en eau, selon le cas	S/O – nouveau champ	Facultatif -- ne s'applique pas à tous les échantillons.
--	Pays/région d'origine/de production	Menu <ul style="list-style-type: none"> • Pas d'informations • Pays (A-Z) 	Facultatif	Nom du pays d'origine ou de production pour les produits finis, se référer au pays d'origine mentionné sur l'étiquette	S/O – nouveau champ	L'information peut ne pas être disponible
--	LM dans le pays/la région d'échantillonnage	Menu: <ul style="list-style-type: none"> • Oui • Non • Pas d'informations 	Obligatoire	Une valeur numérique ou un lien vers la réglementation peut être ajouté facultativement dans Remarques	S/O – nouveau champ	La responsabilité de connaître l'existence de LM dans le pays d'échantillonnage peut incomber à la personne qui saisit les données. Cette information permettra au GTE de savoir si les réglementations nationales ou régionales ont eu une incidence sur les taux de contaminants.
--	Type de produit	Menu: <ul style="list-style-type: none"> • Destiné à une transformation ultérieure (DFP) • Prêt à consommer (RTE) • Sans objet 	Facultatif	DFP et RTE sont définis dans le document CODEX STAN 193-1995.	S/O – nouveau champ	Facultatif car ne s'applique pas à la plupart des échantillons.

		<ul style="list-style-type: none"> • Pas d'informations 				
--	Lieu d'échantillonnage dans la chaîne de production	<p>Menu:</p> <ul style="list-style-type: none"> • Pas d'informations • Ferme • Transport en vrac • Enlèvement pour importation • Industrie • Vente en gros • Vente au détail • Autre 	Obligatoire		S/O – nouveau champ	Le champ peut être obligatoire avec les options Pas d'informations et Autre champ
--	Méthode d'analyse (principe/approche de la méthode d'analyse)	<p>Menu</p> <ul style="list-style-type: none"> • Méthode A • Méthode B • Méthode Z • Autre • Pas d'informations 	Facultatif		S/O – nouveau champ	<p>Peut fournir des informations précieuses en conjonction avec LOQ/LOD.</p> <p>Le menu déroulant doit offrir des options entre les principes/approches des méthodes d'analyse et non pas une très longue liste de méthodes, spécifiant toutes les variantes possibles de certains principes ou de certaines approches de méthodes d'analyse.</p>

APPENDICE II

Rubrique «collecte et soumission de données, et extraction de données» du document d'orientation

**Avant-projet de lignes directrices pour l'analyse des données en vue de l'établissement de limites maximales (LM)
et de l'amélioration de la collecte des données, dans sa version actuelle.**

(Pour information uniquement)

TABLE DES MATIÈRES

Préambule	par. 1-7
Collecte et soumission de données	par. 8-16
Compléter le modèle de GEMS/Aliments	par. 17-32
Extraction de données	par. 33-36

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-dénrée/aliment
 - L'élaboration d'un document de travail qui explore les données préliminaires d'occurrence, les données d'exposition et l'importance mondiale de la combinaison contaminant-dénrée/aliment. **Le document de travail doit également prendre en compte la disponibilité et la qualité des données, afin de permettre une décision éclairée sur d'éventuels nouveaux travaux.**⁹
 - L'accord du CCCF pour entreprendre de nouveaux travaux, y compris la discussion des termes de référence, 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-dénrée/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 soumettent 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 analystes du CCCF extraient les données de la base de données GEMS/Aliments pour élaborer des propositions de LM. Des données externes, issues par exemple de la littérature scientifique, peuvent être référencées, mais elles ne sont généralement pas utilisées pour l'établissement de LM.
3. Avant de démarrer les travaux sur un document de travail ou un document sur les LM, le CCCF peut établir les termes de référence (TdR) pour le groupe de travail et lancer un appel de données. Comme indiqué dans le Manuel de procédure de la Commission du Codex Alimentarius (21^e édition), les termes de référence 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 le travail est censé être achevé. L'appel de données identifie généralement le contaminant et l'aliment/la dénrée 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 et la limite de détection 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'établissement des LM.
5. La gestion des données est une étape clé du travail d'élaboration des normes, et il est dans l'intérêt de tous de disposer de données de bonne qualité (fiabilité des informations collectées, possibilité d'effectuer des analyses statistiques chaque fois que nécessaire, données qui reflètent une image précise de la contamination des aliments...).
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 dans le cadre du développement/de l'élaboration des LM du Codex.
7. Ces lignes directrices sont destinées 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 \(Système mondial de surveillance de l'environnement \(GEMS\)- Programme de surveillance et d'évaluation de la contamination des aliments\)](#). Le

⁹ REP23/CF16 par. 95 et par. 98 (ix)

processus de soumission (upload) et d'extraction (download) des données commence sur le site web, [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 modèles standard et à des modèles 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 la page d'accueil de GEMS/Aliments ([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* sur la page d'accueil 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 modèle Excel et le chargement du modèle 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 le coordinateur 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 mappage, 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 Secrétariat 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 collecte des échantillons, l'année de production, la limite de détection (LOD) et la limite de quantification (LOQ) 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 le coordinateur de GEMS/Aliments pour retirer ou corriger l'ensemble de données, qui devrait être identifiable grâce à l'identifiant du lot.

Compléter le modèle GEMS/Aliments

17. La fiche du modèle pour les soumissions standard (et non pas de blocs)¹⁰ 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 modèle 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

¹⁰ Voir le document *INSTRUCTIONS FOR ELECTRONIC SUBMISSION OF DATA ON CHEMICALS IN FOOD AND THE DIET* pour des informations sur la soumission de modèles pour le chargement de blocs de données.

disponibles dans le menu déroulant, une option est indiquée dans les *INSTRUCTIONS*).

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 d'aliments (colonne B) et niveau 2: Identifiant OMS de l'aliment (colonne C). Une fois les champs «Identifiant local de l'aliment», «Groupe d'aliments» 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 modèle 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 modèle 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 modèle GEMS/Aliments consiste à saisir les résultats de l'analyse individuelle dans l'onglet 3. «Résultats de l'analyse individuelle». Les champs comprennent l'identificateur 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 modèle 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. Les paragraphes 25 à 31 se trouvant sous le tableau explicatif 1 fournissent des informations supplémentaires sur certains champs pour lesquels des conseils à l'intention des personnes qui soumettent des données peuvent être utiles.

Tableau explicatif 1: Champs actuels du modèle GEMS/Aliments

Colonne	Champ	Type de champ	Obligatoire ou Facultatif	Observations
E	Identifiant local de l'aliment	Texte libre	Obligatoire	Nom donné à l'aliment dans la base de données nationale
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.
G	Pays/Région	Menu déroulant	Facultatif	Correspond aux pays ou régions qui soumettent des données; il ne s'agit pas du pays de production.
H	Contaminant	Menu déroulant	Facultatif	Facultatif lorsque le nom du contaminant est saisi dans la fiche 1
I	Origine de l'aliment	Menu déroulant	Facultatif	<ul style="list-style-type: none"> • Production intérieure • Importation • Origine mixte

				• Pas d'informations
J	Date d'échantillonnage	Texte libre (JJJ)	Obligatoire	
K	Représentativité de l'échantillon/ fiabilité	Menu déroulant	Obligatoire	<ul style="list-style-type: none"> • Échantillonnage aléatoire • Échantillonnage ciblé • Pas d'informations
L	Identification en laboratoire	Texte libre	Facultatif	Laboratoire soumettant les résultats
M	Assurance qualité analytique	Menu déroulant	Facultatif	<ul style="list-style-type: none"> • Assurance qualité interne et normes de référence uniquement. • Participation réussie à des tests d'aptitude pertinents pendant 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
N	Unités de mesure pour les taux de contaminants	Menu déroulant	Obligatoire	<ul style="list-style-type: none"> • mg • ug • ng • pg • Bq
O	LOD	Texte libre	Obligatoire pour les résultats non quantifiés si la LOQ n'est pas fournie	
P	LOQ	Texte libre	Obligatoire pour les résultats non quantifiés si la LOD n'est pas fournie	
Q	Résultats basés sur	Menu déroulant	Obligatoire	<ul style="list-style-type: none"> • Teneur en graisses • Poids sec • Tel quel (cru, frais) • Tel que consommé
R	Portion analysée	Menu déroulant	Obligatoire	<ul style="list-style-type: none"> • Uniquement comestible • Aliment total (comestible + non comestible)
S	État de l'aliment analysé	Menu déroulant	Facultatif	<ul style="list-style-type: none"> • Cuit • Cru • Pas d'informations
T	Résultats	Texte libre	Obligatoire	
U	Données individuelles versus données agrégées	Menu déroulant	Facultatif	<ul style="list-style-type: none"> • Individuelles • Agrégées
V	Confidentialité des données	Menu déroulant	Facultatif	<ul style="list-style-type: none"> • Oui • Vide
W	Remarques/Références	Texte libre	Facultatif	

25. **Identifiant local de l'aliment.** 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.
26. **Unités.** Veillez à ce que l'unité de 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.

27. **Numéro de série.** Un numéro de série (identifiant de l'échantillon) doit être utilisé pour chaque échantillon. Si des informations sur plusieurs contaminants sont soumises pour un même échantillon, le même numéro de série doit être utilisé. (Veuillez noter qu'il est possible de saisir plusieurs contaminants dans un seul modèle.) Les institutions nationales doivent coordonner l'utilisation du même numéro de série pour toutes les soumissions d'un même échantillon.
28. **Pays/région.** Ce champ indique les pays ou régions qui soumettent des données; il ne s'agit pas du pays de production ni du pays d'origine.
29. **Données agrégées.** 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 LM appropriées, qui sont basées sur l'observation de la distribution des données et des centiles supérieurs dépassant les limites maximales (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. Le coordinateur de GEMS/Aliments ou un statisticien doit être consulté avant l'inclusion de données agrégées. Si des données agrégées sont incluses dans une analyse de LM, cela doit être indiqué dans le document destiné au GTE.
30. **Données confidentielles.** Les pays peuvent soumettre des données sous la mention «confidentiel» 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.
31. **Remarques/Références** Ce champ est utilisé pour noter des remarques et/ou des références relatives aux données. En règle générale, les personnes qui soumettent des données utilisent cette colonne pour ajouter des informations qui ne sont pas saisies dans les champs du modèle. À titre d'exemple d'informations incluses dans cette colonne par le passé, citons les informations figurant sur les étiquettes des produits (telles que les ingrédients principaux ou les noms détaillé des produits), les informations relatives à la composition (telles que le pourcentage de cacao), le pays d'origine ou de production, la méthode utilisée pour l'analyse, etc. Une référence à un appel de données spécifique est une autre information qui peut être inscrite dans cette colonne.
32. **Erreurs.** Avant d'introduire des données, la personne qui les soumet doit relire le fichier attentivement afin de détecter 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 modèle. Des ensembles de données peuvent être rejetés pour diverses raisons, dont certaines sont énumérées ci-dessous. Le coordinateur de GEMS peut être contacté pour obtenir de l'aide.
 - a. Résultat communiqué < LOD, LOQ ou LOD manquante lorsque le résultat est ND, LOD communiqué > LOQ
 - 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

33. 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.
34. Afin d'identifier l'ensemble de données le plus précis à extraire pour l'élaboration de propositions de LM, il est préférable de consulter le coordinateur de 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 du coordinateur 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.

35. 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.
36. Il est important de conserver un enregistrement de tous les filtres et termes de recherche pour le rapport du GTE.

APPENDICE III**Une liste des titres/sujets à aborder dans les rubriques «sélection/nettoyage des données» et «analyse des données»****(pour examen)****Sélection/nettoyage de données**

Considérations générales

Manque d'informations sur les données fournies

Sélection et nettoyage – Traitement de données

- avec informations manquantes
- avec informations incorrectes supposées
- provenant d'échantillons soupçonnés d'être frauduleux ou adultérés
- provenant d'un échantillonnage ciblé
- valeurs aberrantes/extrêmes (y compris les méthodes de détermination des valeurs aberrantes/extrêmes)
- Considérations relatives à la limite de quantification (LOQ) et à la limite de détection (LOD)

Analyse de données: génération d'un aperçu des données

- Aperçu présentant quels pays, combien de points de données, quelles années, la période de couverture des données
- Décision sur la couverture géographique des données d'occurrence fournies (y compris la possibilité de combiner différents ensembles de données ou de les garder séparés)
- Décision sur la période couverte par les données d'occurrence fournies

Analyse statistique des données d'occurrence/ traitement d'ensembles de données pour l'élaboration de LM

Considérations générales

Nombre suffisant d'échantillons

- nombre minimum d'échantillons pour l'estimation des valeurs de centile élevées

Traitement d'ensembles de données

- avec un faible nombre de points de données
- dont les données concernant des aliments individuels sont insuffisantes, mais dont les données concernant le groupe d'aliments sont suffisantes
- avec une large proportion de données censurées à gauche (y compris l'utilisation de méthodes de substitution)

Effectuer une analyse statistique

- Génération de diagrammes/graphiques et de courbes se rapportant à la distribution des données d'occurrence
- Agrégation des données et calcul des statistiques descriptives

Calcul des taux de rejet aux LM hypothétiques

- Estimation des LM hypothétiques
- Calcul des taux de rejet aux LM hypothétiques
- Évaluation de l'impact d'une LM sur le taux de rejet
- Amélioration du calcul des taux de rejet

Calcul des effets des LM sur la réduction de l'exposition alimentaire à des LM hypothétiques

- Calcul de l'exposition alimentaire et de la réduction à des LM hypothétiques
- Évaluation de l'impact d'une LM sur l'exposition alimentaire

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

Présentation des données dans les rapports des GTE au CCCF

ANNEXE: GLOSSAIRE DES TERMES UTILISÉS

Sujets de discussion supplémentaires (après la 17^e session du CCCF)

- 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)).
- Des lignes directrices supplémentaires restent 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 la denrée ait été produite 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 une denrée 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)).
- 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 (lignes directrices relatives 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).

APPENDICE IV

Rubriques «Sélection/nettoyage des données» et «analyse des données» du projet de Lignes directrices sur l'analyse des données pour le développement des limites maximales (LM) et pour l'amélioration de la collecte des données, dans lequel les résultats des discussions du CCCF16 ont été intégrés avec un glossaire de termes.

(Pour information uniquement)

TABLE DES MATIÈRES

Sélection/nettoyage de données

Questions générales	par. 37-41
Manque d'informations sur les données fournies	par. 42-47
Traitement des données pour lesquelles une erreur de déclaration est supposée	par. 48-50
Données provenant d'échantillons soupçonnés d'être frauduleux ou falsifiés	par. 51
Données provenant d'un échantillonnage ciblé	par. 52-53
Limite de quantification (LOQ) et limite de détection (LOD)	par. 54-58
Analyse de données: génération d'un aperçu des données	
Aperçu présentant quels pays, combien de points de données, quelles années, période de couverture des données	par. 59
Décision sur la couverture géographique des données d'occurrence fournies	par. 60-65
Décision sur la période couverte par les données d'occurrence fournies	par. 66-71

Analyse statistique des données d'occurrence pour l'élaboration de LM

Considérations générales	par. 72-75
Nombre suffisant d'échantillons	
Nombre minimum d'échantillons pour l'estimation des valeurs de centile élevées	par. 76-80
Traitement des ensembles de données avec un faible nombre de points de données	par. 81-85
Le traitement d'ensembles de données concernant des denrées alimentaires individuelles est insuffisant les données pour le groupe alimentaire sont suffisantes	par. 86-89
Traitement d'ensembles de données avec une large proportion de données censurées à gauche	par. 90-92
Méthodes de substitution	par. 93-98
Traitement des ensembles de données multiples	par. 99-102
Cas où les ensembles de données peuvent être combinés	par. 103-105
Cas où les ensembles de données individuels peuvent être utilisés	par. 106-110
Détermination et traitement des valeurs aberrantes/extrêmes	par. 111-112
Avant la détermination des valeurs aberrantes	par. 113-114
Test statistique des valeurs aberrantes	par. 115-116
Autres méthodes d'identification des valeurs aberrantes	par. 117
Décision sur le traitement des valeurs aberrantes possibles	par. 118-120
Réalisation d'une analyse statistique	
Génération de diagrammes/graphiques et de courbes se rapportant à la distribution des données d'occurrence	par. 121-127
Agrégation des données et calcul des statistiques descriptives	par. 128-130
Calcul des taux de rejet à des LM hypothétiques	
Estimation des LM hypothétiques	par. 131-134
Calcul des taux de rejet à des LM hypothétiques	par. 135-136

Évaluation de l'impact d'une LM sur le taux de rejet	par. 137-138
Amélioration du calcul des taux de rejet	par. 139-141
Calcul des effets des LM sur la réduction de l'exposition alimentaire à des LM hypothétiques	
Calcul de l'exposition alimentaire et de la réduction à des LM hypothétiques	par. 142-152
Évaluation de l'impact d'une LM sur l'exposition alimentaire	par. 153-154
Amélioration du calcul des taux de réduction de l'exposition	par. 155-157
Présentation des données dans les rapports des GTE au CCCF	par. 158
Présentation de l'analyse des données: analyse statistique	par. 159-167
APPENDICE: Glossaire des termes	

SELECTION/NETTOYAGE DE DONNÉES

Généralités

37. Pour le nettoyage des données, il est recommandé de faire appel à un expert du contaminant en question, afin qu'il puisse déterminer si les schémas des données sont irréguliers ou non.
38. Toutes les étapes du nettoyage des données doivent être enregistrées et décrites dans le document final, par exemple les raisons des exclusions, le nombre d'exclusions à chaque étape du nettoyage, etc. Dans tous les cas, il convient de conserver les données exclues et les détails concernant les données exclues (d'une région spécifique, d'une année spécifique, d'un fournisseur de données spécifique,...).
39. Une analyse de sensibilité sur l'impact de l'exclusion des données peut être réalisée pour déterminer l'impact de l'exclusion des données.
40. Le nettoyage ne concerne que l'ensemble de données extraites, tandis que 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.
41. Pour le développement des LM, seules les données de la base de données GEMS/Aliments seront utilisées. Les données non GEMS/Aliments ne peuvent être utilisées qu'à des fins d'analyse complémentaire, lorsque les données disponibles dans la base de données GEMS/Aliments sont limitées ou lorsque les données ne sont pas disponibles dans la base de données GEMS/Aliments pour certaines périodes ou régions, en particulier pour les pays producteurs primaires.

Lorsqu'il s'agit de données directement soumises à l'EWG par le(s) pays ou le(s) observateur(s) ou obtenues par la recherche documentaire sans passer par la base de données GEMS/food, ces données sont également soumises à la procédure de nettoyage, le cas échéant.

Manque d'informations sur les données fournies

42. Si tous les champs obligatoires sont remplis (voir la rubrique «Collecte et transmission des données») et que les données peuvent être téléchargées dans la base de données GEMS/Aliments, les données ne doivent pas être exclues.
43. Dans certains cas, les données de la base de données GEMS/Aliments ne fournissent pas toutes les informations qui seraient utiles au GTE pour réaliser une analyse (par exemple, il n'est pas clair si l'aliment est séché ou frais). En cas d'informations manquantes, le point de contact du pays ou de l'organisation qui soumet les données doit être contacté dans un premier temps pour permettre l'obtention d'un ensemble de données complet. Le point de contact du pays ou de l'organisation qui soumet les données peut avoir besoin de contacter les personnes impliquées dans le processus d'élaboration des données, comme l'échantillonnage, l'analyse chimique, l'analyse des données, afin d'identifier toute information manquante, lorsqu'il est contacté par le GTE.
44. Si des informations manquantes sont fournies par le pays ou l'organisation qui soumet les données, l'administrateur du GEMS/Aliments doit également être informé par la personne qui soumet les données afin que les informations fournies soient incluses dans la base de données du GEMS/Aliments, et pas seulement dans l'ensemble de données en cours d'analyse par la présidence du GTE.
45. Deuxièmement, il convient de déterminer dans quelle mesure les informations manquantes affectent l'analyse des données.
Il ne faut pas fixer de règles générales susceptibles d'entraîner des exclusions inutiles: les données comportant des informations manquantes peuvent encore être utiles. Le même niveau de détail (c'est-à-dire le même volume d'informations) concernant les échantillons peut ne pas être nécessaire pour le développement/l'élaboration de toutes les limites maximales. Par exemple, certains produits de base tels que les boissons peuvent ne pas nécessiter le même niveau de détail que celui des céréales, comme l'étape de la transformation, afin de proposer des limites maximales. En outre, dans certains cas, les informations manquantes peuvent être déduites d'autres informations fournies. Par exemple, si l'échantillon est décrit comme du paprika séché, il est évident que l'état de la denrée alimentaire analysée est «séché», même si le champ S de la base de données GEMS/Aliments n'est pas complété.
46. Exemples d'informations manquantes pour lesquelles les données devraient éventuellement être exclues d'une analyse plus approfondie:
 - toutes les données d'un ensemble de données sont déclarées < LOQ et la LOQ n'est pas fournie. Étant donné que les informations relatives à la LOQ doivent obligatoirement être téléchargées dans la base de données GEMS/Aliments si le résultat analytique est < LOQ, cette situation peut ne pas se produire lorsque toutes les données ont été extraites de la base de données GEMS/Aliments, mais elle peut se

produire si l'on considère les données directement soumises au GTE sans passer par la base de données GEMS/Aliments.

- l'unité dans laquelle le résultat est rapporté ou la base sur laquelle le résultat est exprimé est manquante.
- l'état de l'aliment échantillonné (par exemple, s'il est séché ou frais) n'est pas indiqué.
- description adéquate du produit (par exemple, l'analyse porte sur le «maquereau», mais le produit est décrit comme étant du «poisson»)

47. Exemples d'informations manquantes, mais les données pourraient quand même être utilisées pour une analyse plus approfondie (ceci doit être évalué au cas par cas, car pour certaines combinaisons de contaminants alimentaires, les informations ci-dessous pourraient être considérées comme nécessaires et les informations manquantes pourraient donc constituer une base d'exclusion):

- informations sur l'échantillonnage: type d'échantillonnage, année d'échantillonnage, lieu d'échantillonnage...
- l'état du produit, par exemple frais ou séché
- la méthode d'analyse utilisée, ses données de validation et ses caractéristiques de performance (telles que la récupération, l'incertitude, la LOQ
- une cartographie alimentaire incorrecte a été utilisée pour décrire le produit
- Code d'usages utilisé ou non, années écoulées depuis sa mise en œuvre
- lorsque la LM est fixée ou envisagée pour la somme des composants et que les données sont déclarées non pas pour tous les composants, mais pour celui ou ceux qui contribuent de manière significative à la somme

Traitement des données pour lesquelles on peut raisonnablement supposer que l'unité des données fournies ou la base sur laquelle les données sont déclarées (par exemple, base grasse contre poids entier) n'est pas correcte (parties pertinentes des par. 111 - 120 - Détermination des valeurs aberrantes/extrêmes et leur traitement à intégrer dans la présente partie).

48. S'il existe des indications claires que l'unité dans laquelle les données sont exprimées est incorrecte ou que la base sur laquelle les données sont exprimées est incorrecte, le point de contact du pays qui a soumis les données peut être contacté pour les corrections. Les données ne peuvent être remplacées par une unité corrigée que si l'auteur des données est d'accord. Si une erreur ne peut être confirmée, que des corrections ne peuvent être apportées et que ces échantillons sont effectivement déclarés de manière incorrecte, ces données doivent être exclues de l'analyse ultérieure des données. Les échantillons contenant des informations incorrectes et leurs valeurs corrigées doivent être enregistrés et présentés dans le document final.

49. Exemples d'«indications claires» permettant de contacter les auteurs pour qu'ils apportent d'éventuelles corrections et qu'ils soumettent à nouveau leurs documents:

- Les niveaux dans un ensemble de données de 200 résultats sont compris entre 0 et 20. Toutes les données sont exprimées en µg/kg, sauf 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, elles sont identifiées comme d'éventuelles valeurs aberrantes (voir paragraphe 117).
- Limites provenant d'un aliment ayant une teneur en matières grasses typique de 5 % dans un ensemble de données de 200 résultats dont toutes les données sont désignées comme étant exprimées sur la base du poids total. 195 résultats 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 peut-être qu'ils ont été rapportés sur la base de la matière grasse 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 paragraphe 117).

50. 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 grains de riz vs riz décortiqué vs riz poli), des arguments similaires s'appliquent comme dans la rubrique B2. 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. Toute hypothèse de ce type doit être enregistrée et présentée dans le document final. Si les informations sont pertinentes pour l'analyse des données

et qu'une hypothèse raisonnable ne peut être formulée, ces données doivent être exclues de l'analyse ultérieure des données, à moins que les informations nécessaires ne soient obtenues.

Données provenant d'échantillons soupçonnés d'être frauduleux/économiquement falsifiés (la partie pertinente des paragraphes 111 - 120 - Détermination des valeurs aberrantes/extrêmes et traitement de ces valeurs doit être intégrée dans cette partie)

51. Les données qui sont clairement liées à des échantillons frauduleux ou économiquement falsifiés, sur la base des informations pertinentes fournies par l'auteur des données, doivent être exclues de l'analyse et l'exclusion doit être documentée. Les signes possibles d'échantillons frauduleux/économiquement falsifiés sont les suivants:
- certains échantillons sont d'un ordre de grandeur plus élevé que d'autres (par exemple, 1 contre 15 µg/kg) ou
 - la variabilité temporelle des données, par exemple, les données sont beaucoup plus élevées pour une année de l'ensemble de données.

Toutefois, ce contraste peut également résulter de la variabilité naturelle (par exemple, un niveau élevé de mycotoxines dû à des conditions climatiques spécifiques dans une certaine région), de sorte que la nature du contaminant doit être prise en compte lors de l'évaluation des données.

Données provenant d'un échantillonnage ciblé (les parties pertinentes des paragraphes 111 - 120 - Détermination des valeurs aberrantes/extrêmes et leur traitement doivent être intégrées dans cette partie)

52. L'échantillonnage ciblé diffère de l'échantillonnage aléatoire en ce sens qu'il s'agit d'une stratégie d'échantillonnage distincte visant des lots spécifiques. En principe, ces données ne devraient pas être utilisées dans la dérivation des LM, car les données provenant d'un échantillonnage ciblé ne reflètent pas les niveaux réalisables dans des situations normales.
53. Il convient de noter que même dans le cas d'un échantillonnage aléatoire, un certain biais peut être introduit, car il peut y avoir des raisons d'échantillonner plus largement 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.

Nouvelle partie: Valeurs aberrantes/extrêmes (y compris les méthodes de détermination des valeurs aberrantes/extrêmes) (les parties pertinentes des paragraphes 111 à 120 doivent être incluses ici)

Considérations relatives à la limite de quantification (LQ) et à la limite de détection (LOD) (voir également les paragraphes 90-92)

54. Il convient de noter que les différentes méthodes d'analyse fournissent des limites de détection et des limites de quantification différentes. Une LOQ élevée ne signifie pas automatiquement que les données doivent être exclues. L'adéquation de la LOQ de l'ensemble de données pouvant être utilisé pour la dérivation des LM doit être évaluée par rapport à la LM proposée, comme décrit ci-dessous.
- Lorsqu'aucune LOQ/LOD n'est fournie pour un ensemble de données spécifique:
 - L'ensemble de données contient (presque) tous les résultats quantifiés.
 - L'ensemble de données contient une part importante de données censurées à gauche (c'est-à-dire < LOQ) et aucune LOQ/LOD n'est fournie.
 - Lorsque la LOQ est fournie pour un ensemble de données spécifique:
 - Ensemble de données dont la LOQ est nettement inférieure à la LM considérée.
 - Ensemble de données dont la LOQ se situe dans la fourchette de la LM considérée
 - Ensemble de données dont la limite de quantification est supérieure à la limite de quantification proposée à l'étude
55. Lignes directrices pour les scénarios susmentionnés (paragraphe 54)
- Dans le cas où aucune LOQ/LOD n'est fournie pour un ensemble de données spécifique,
 - le pays demandeur pourrait être contacté dans un premier temps pour permettre la fourniture de ces informations (c'est-à-dire la limite de détection et/ou la limite de qualité).
 - Dans le cas où l'ensemble de données contient (presque) tous les résultats quantifiés: l'ensemble

de données peut être utilisé.

- Si l'ensemble de données contient une part importante de données censurées à gauche, l'ensemble de données ne doit pas être utilisé.
- Dans le cas où la LOQ est fournie:
- Niveau seuil à déterminer pour la LOQ (exemples: $LOQ < LM$ en cours de discussion, $LM < 0,5 LM$ en cours de discussion).
56. Si la quasi-totalité des données de l'ensemble de données sont inférieures à la LOQ/rapportées comme non détectées (ND), il n'est pas possible d'estimer des valeurs de centile élevées pour établir les taux de rejet. Lorsqu'il y a de nombreuses données <LOQ et un plus petit nombre de valeurs quantitatives, l'ensemble de données doit être traité au cas par cas en suivant les lignes directrices indiquées aux paragraphes 81 à 98. Il n'est pas approprié de calculer des valeurs de centile élevées en utilisant uniquement les valeurs quantitatives, ce qui peut entraîner des propositions inutilement élevées pour les LM.
57. Il convient de définir des critères indiquant quand certaines données doivent être exclues de l'ensemble de données en raison d'une limite de détection (LOD) inadéquate (par exemple, la limite de détection est supérieure à la limite de quantification proposée, la limite de détection est supérieure de «x» ordres de grandeur à la limite de détection la plus basse de l'ensemble de données) ou si l'ensemble de données doit être exclu de l'analyse, étant donné que la suppression de données individuelles peut introduire un biais.
58. Niveaux de contaminants qui sont une somme de composants et pour lesquels certains composants sont inférieurs à la LOQ.
- La règle générale est que les niveaux de contaminants qui sont une somme de composants sont rapportés en tant que limite inférieure, c'est-à-dire que les composants non quantifiés sont mis à 0. Des informations sur la limite de détection ou la limite de quantification des différents composants de la somme doivent être fournies. Toutefois, dans le cas de tous les cas de non-détection, on pourrait considérer qu'il n'y a pas lieu d'exclure les données si la limite de quantification des différents composants n'est pas indiquée, étant donné que le résultat de la somme sera égal à 0, quelle que soit la limite de quantification.
 - Lorsque seules des données sur des composants individuels sont communiquées sans résultat total, les données individuelles peuvent être additionnées pour obtenir un résultat global: dans ce cas, la limite de détection ou la limite de quantification doit être indiquée.
 - Dans certains cas, il peut être approprié de déclarer les niveaux de contaminants qui sont une somme de composants en utilisant une approche de limite moyenne ou de limite supérieure; cependant, ces cas doivent être clairement identifiés à l'avance avant la soumission des données. Dans ce cas, des limites de détection (LOD) ou des limites de qualité (LOQ) pour les données des différents composants sont nécessaires.

Analyse des données: génération d'un aperçu de données

Vue d'ensemble des pays, du nombre de points de données, des années, de la période de couverture des données

59. 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 d'origine, l'année de production, la quantité de données incluses et exclues) doit être fourni dans un tableau. Toutes les mesures prises dans le cadre du nettoyage, la justification et les hypothèses formulées doivent être fournies avec la vue d'ensemble. En outre, il pourrait être utile de fournir des informations (par exemple, de la FAO) sur les principales régions de production de la denrée en question. Sur la base de cette vue d'ensemble, il est possible de procéder à une sélection plus poussée de la couverture géographique et de la période concernées.

Décision sur la couverture géographique des données d'occurrence fournies, y compris l'examen de la possibilité de combiner ou de garder séparés différents ensembles de données (les parties pertinentes des paragraphes 99-110 sur le traitement d'ensembles de données multiples doivent être intégrées dans cette partie)

60. Les pays qui soumettent des données à GEMS/Aliments doivent s'assurer que les données soumises sont aussi représentatives que possible au niveau national.
61. Il doit y avoir au minimum une représentation des régions de production importantes pour le commerce international. Il est donc important que l'origine de la denrée alimentaire soit indiquée dans la base de données GEMS/Aliments (voir la rubrique «Collecte et transmission des données»). Dans ce contexte, les données des régions productrices doivent être considérées par rapport aux données des pays qui importent les denrées alimentaires, car ces dernières peuvent être faussées si les denrées alimentaires doivent se conformer aux

exigences du pays importateur, comme une LM déjà établie dans ce pays. 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. En effet, au cours du transport depuis le pays producteur, une contamination supplémentaire a pu avoir lieu (par exemple, production de mycotoxines).

62. 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, mais dans ce cas, des garanties devraient être fournies que les ensembles de données des pays producteurs reflètent la mise en œuvre des bonnes pratiques prévues dans les codes d'usages du Codex et sont représentatifs des produits qui feraient l'objet d'un commerce international.
63. Si possible, une analyse de sensibilité sur l'utilisation des données des pays producteurs par rapport aux pays importateurs pourrait être réalisée pour guider la sélection des données.
64. Comme les LM du Codex sont des normes mondiales, une approche par défaut du traitement des données consiste à analyser les données par région. Ce n'est que s'il y a suffisamment de données indiquant des différences importantes dans les niveaux déclarés entre les pays d'une région que l'analyse peut être effectuée par pays. Il convient de noter que, dans le cas d'une approche par pays, cette démarche doit être effectuée pour les principaux pays producteurs de la région et que des données suffisantes doivent être disponibles. Le nombre de points de données considérés comme suffisants doit être examiné au cas par cas.
65. Lignes directrices pour les ensembles de données qui manquent de couverture géographique:
 - Si la ou les régions pour lesquelles les données manquent sont des régions de production importantes et à condition que les régions productrices s'engagent clairement, quelques années supplémentaires sont accordées pour la collecte de données avant de poursuivre la discussion sur les propositions de LM. Après l'expiration des années supplémentaires accordées, la discussion sur les LM se poursuit sur la base des données disponibles, que la couverture géographique ait été atteinte ou non.
 - Si la/les région(s) pour laquelle/lesquelles les données sont insuffisantes n'est/sont pas une/des région(s) de production importante(s): la discussion sur les LM sera poursuivie sur la base des données disponibles.
 - Si la ou les régions productrices importantes ne s'engagent pas à fournir les données supplémentaires, la discussion sur les LM doit se poursuivre sur la base des données disponibles ou il peut être décidé d'interrompre la discussion sur une LM.

Décision sur la période couverte par les données d'occurrence fournies (les parties pertinentes des paragraphes 99-110 sur le traitement des ensembles de données multiples doivent être intégrées dans cette partie)

66. Il convient que les données d'occurrence fournies se rapportent à plusieurs années de production pour le développement des LM (elles peuvent être différentes pour différents types de contaminants: mycotoxines, toxines végétales, biotoxines marines, contaminants de transformation, contaminants environnementaux, en fonction de la variation annuelle supposée ou de l'évolution de la contamination dans le temps).
67. 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, les variations annuelles sont moins importantes et des données plus récentes 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.
68. En outre, il pourrait être utile d'étudier/inclure des données plus anciennes pour savoir si certaines espèces/sous-groupes d'un groupe ont tendance à présenter des niveaux plus élevés.
69. En outre, il pourrait être utile dans certains cas 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.
70. Dans le cas où 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, sauf si un pays indique que les bonnes pratiques avaient déjà été mises en œuvre avant l'établissement du Code.

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 niveaux entre pays voisins au sein d'une même région qui ne peuvent être expliquées par des facteurs géographiques.

71. Il est souvent difficile de juger, à partir des ensembles de données eux-mêmes, 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 si le pays qui soumet ces données a déjà mis en place ses propres mécanismes de surveillance. Si le GTE exclut des données sur la base de la non-application d'un Code d'usages, les exclusions et leur justification doivent être clairement documentées dans le document du GT.

ANALYSE STATISTIQUE DES DONNÉES D'OCCURRENCE /TRAITEMENT D'ENSEMBLES DE DONNÉES POUR LE DEVELOPPEMENT DES LM

CONSIDÉRATIONS GÉNÉRALES

72. Pour le développement des LM, la qualité globale des données est essentielle. Les données d'occurrence, obtenues idéalement par échantillonnage statistique (Réf. CXG 50-2004 Directives générales sur l'échantillonnage), et l'analyse à l'aide de méthodes validées avec des limites de quantification et des limites de détection appropriées dans des laboratoires disposant de systèmes d'assurance qualité. Les données obtenues doivent être soigneusement examinées et extraites/nettoyées. Les données extraites/nettoyées doivent faire l'objet d'une analyse statistique. Toutefois, pour élaborer une LM appropriée, il convient de tenir compte non seulement des résultats de l'analyse statistique, mais aussi de tout risque sanitaire associé au contaminant/à la toxine en question (toxicité et exposition alimentaire, seule ou combinée).
73. Pour décider de la méthode statistique à utiliser, il convient d'examiner attentivement le modèle de distribution de l'ensemble de données. En général, la distribution des données sur les contaminants dans les aliments tend à être asymétrique avec une longue queue à droite, par exemple une distribution log-normale. Pour de telles distributions, l'utilisation de méthodes statistiques paramétriques, basées sur la distribution normale, n'est pas appropriée.
74. La Norme générale pour les contaminants et les toxines présents dans l'alimentation humaine et animale (CXS 193-1995) (ci-après dénommée 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 l'alimentation humaine et animale*» Cela signifie que pour développer une LM, il est nécessaire d'estimer/déterminer des valeurs de centile élevées (généralement des valeurs de 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 la valeur arrondie de la valeur de centile) et le pourcentage d'échantillons qui dépassent la LM proposée lorsque la LM est identique à la valeur du 98^e centile.

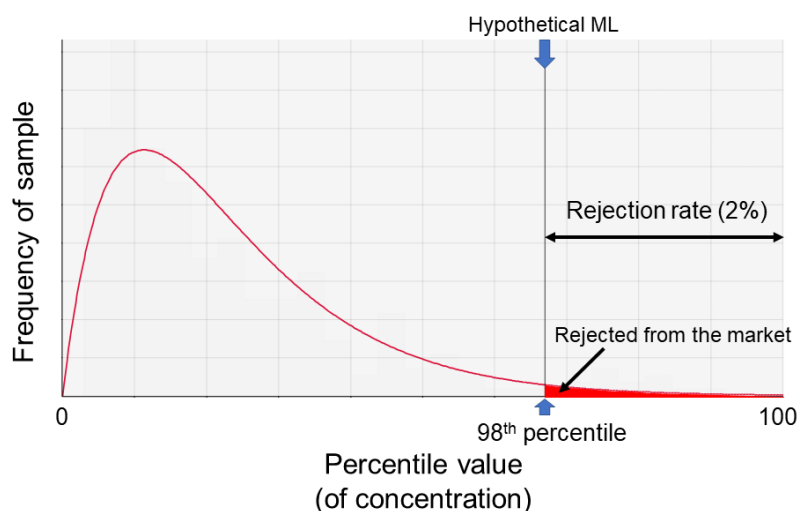


Figure 1. Représentation simplifiée de la relation entre une valeur de centile élevée, une LM hypothétique, le taux de rejet et le pourcentage d'échantillons dépassant la LM proposée.

Remarque: Dans ce qui précède, on suppose que la LM hypothétique est identique à la valeur du 98^e centile.

75. Les rubriques suivantes expliquent les considérations à prendre en compte avant de procéder à une analyse statistique 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.

Nombre suffisant d'échantillons

La fiabilité des valeurs estimées de centile élevées dépend du nombre de points de données disponibles pour le calcul. Les valeurs de centile calculées sur un petit nombre de données peuvent ne pas être statistiquement solides.

Bien qu'il existe un certain nombre de lignes directrices sur le nombre minimum de points de données nécessaires pour calculer une valeur de centile élevée donnée, aucune n'a été officiellement utilisée par le CCCF. Le nombre minimum d'échantillons est l'un des nombreux facteurs importants à prendre en compte lors de la conception des enquêtes (par exemple, les TdR et l'appel de données) afin d'améliorer la collecte des données.

La NGCTAHA stipule ce qui suit:

«Quand il apparaît que les modèles de contamination sont suffisamment bien compris et seront similaires à l'échelle mondiale, des données moins approfondies pourront éventuellement suffire»;

«Des limites maximales peuvent être établies pour des groupes de produits, quand on dispose d'informations suffisantes sur le modèle de contamination affectant l'ensemble du groupe, ou quand il existe d'autres arguments en faveur de l'extrapolation»; et

«Des limites maximales devraient être établies uniquement pour les contaminants qui présentent un risque sérieux pour la santé publique et qui posent ou peuvent poser un problème dans le domaine du commerce international et uniquement pour les aliments dans lesquels un contaminant peut être présent en quantité suffisamment importante pour affecter l'exposition totale du consommateur».

Nombre minimum d'échantillons pour l'estimation des valeurs de centile élevées

76. Pour développer une LM, il est nécessaire d'estimer les valeurs de centile élevées (généralement les valeurs du 95^e centile) d'un ensemble de données, comme indiqué au paragraphe 74. La possibilité d'estimer ces valeurs de centile élevées avec un niveau de confiance élevé dépend du nombre d'échantillons. Il est donc important de vérifier si le nombre d'échantillons (ou de données) est suffisant pour estimer des valeurs de centile élevées. En général, le nombre minimum d'échantillons doit être déterminé au moment de la conception des études sur les contaminants.
77. Actuellement, trois options sont disponibles pour calculer le nombre minimum d'échantillons 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 d'échantillons est obtenu par 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 d'échantillons).
 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 NC supérieur à 0,95.
 - **Option 2:** Calcul basé sur la règle de Kroes *et al.* (2002)
 - ²¹ Le nombre minimum d'échantillons 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 d'échantillons).
 Aucune information n'est disponible dans la référence sur les niveaux de confiance.
 - **Option 3:** Calcul basé sur les descriptions de Conover (1971)²² 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 d'échantillons est obtenu par 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

²¹ Kroes, Robert, et al. «Assessment of intake from the diet.» Food and Chemical Toxicology 40.2-3 (2002): 327-385.

²² Conover WJ, 1971. Practical nonparametric statistics. Wiley, New-York, États-Unis.

d'échantillons; et $x_{1-\alpha}$, $(1-\alpha)$ 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 ($\alpha=0,01$)

78. La formule utilisant le même concept que l'option 1 (en remplaçant le centile/100 par $(1-\text{taux de violation}/100)$) a été utilisée pour les tests de conformité développés par le CCPR (CXG 33-1999) et le CCRVDF (CXG 71-2009). Certaines lignes directrices²³²⁴ 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 d'échantillons 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 %.

Tableau 1. Nombre minimum d'échantillons pour obtenir des valeurs de centile élevées.

Option	Nombre minimum d'échantillons pour obtenir les valeurs de centile suivantes				
	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 fournie*)	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.*

79. 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 d'échantillons par rapport aux autres options. Il s'agit de l'option la plus réalisable et la plus utilisée lors de l'élaboration précédente des LM par le CCCF, compte tenu du nombre de données dont disposent les GTE.
80. Le tableau 1 sert de guide pour comprendre le nombre minimum d'échantillons nécessaires pour l'ensemble de données qui sera utilisé pour estimer 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 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 ou organisation(s) membre(s) contienne un nombre de points de données supérieur au 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 d'échantillons est requis. Pour le traitement d'ensembles de données multiples et la décision de les combiner ou non, voir les paragraphes 99 à 110.

La 16^e session du CCCF a accepté un nombre minimum **provisoire** de 59 échantillons pour une estimation du 95^e centile avec un niveau de confiance de 95 % (option 1 dans le Tableau 1) (REP23/CF16 paragraphes 93, 94, 98 (vi) (a)). Cependant, la 16^e session du CCCF a également convenu que le nombre minimum d'échantillons pour estimer les valeurs de centile élevées avec un niveau de confiance élevé doit être discuté plus avant (REP23/CF16 paragraphes 93 et 98 (vi) (a))

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

81. Lorsque le risque associé est important au point qu'il est jugé nécessaire d'établir une LM, une taille d'échantillon inférieure à celle spécifiée dans le tableau 1 serait toujours considérée comme adéquate tant que le niveau de confiance des valeurs de centile élevées estimées n'est que légèrement inférieur au niveau de confiance élevé attendu, par exemple 95 %. Par exemple, le niveau de confiance pour l'option 1 peut être obtenu à partir 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 d'échantillons)

82. Si la taille de l'échantillon collecté est insuffisante pour développer la LM, des appels de données

²³ 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

²⁴ EFSA Journal, 2014. Guidance on the EU Menu methodology. 12(12):394

supplémentaires pourraient être lancés si nécessaire (voir Collecte/soumission de données sur les événements). Toutefois, si, après plusieurs appels de données, le nombre d'échantillons disponibles reste très inférieur au nombre minimal requis, il convient de décider, au cas par cas, de développer une LM en utilisant l'ensemble de données limité ou d'interrompre les travaux en fonction du niveau de risque (toxicité, ingestion alimentaire, etc.). Les LM devraient être établies même si le nombre d'échantillons disponibles est faible, lorsqu'une LM est nécessaire de toute urgence du point de vue de la protection de la santé des consommateurs. Si des données suffisantes sont disponibles à l'avenir, une révision des LM précédemment établies peut être envisagée.

83. Pour les produits qui ne sont pas consommés couramment et/ou qui ne font pas l'objet d'un commerce international, les données disponibles sur l'occurrence peuvent être insuffisantes. Le GTE chargé d'établir les LM devrait envisager de recommander au CCCF que la LM demandée pour la combinaison denrée/contaminant ne réponde pas aux critères décrits dans la NGCTAHA et le Manuel de procédure («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'établissement des LM.
84. Si le nombre de données est nettement inférieur au nombre requis pour chaque option du tableau 1 et qu'il n'y a pas de raison sérieuse de développer immédiatement une LM, il n'est pas nécessaire d'effectuer d'autres analyses statistiques. Des appels de données supplémentaires peuvent être nécessaires pour établir une LM statistiquement solide, ou le travail devrait être reporté jusqu'à ce que davantage de données soient fournies.
85. Lors de la révision des LM existantes, même si seules quelques données provenant de régions limitées sont disponibles et qu'il est probable qu'aucune nouvelle donnée ne sera générée, les LM ne doivent pas être automatiquement révoquées en raison de la petite taille de l'échantillon, à moins que la valeur de la LM ne soit incompatible avec les bonnes pratiques actuelles ou les données toxicologiques actuelles. S'il existe un risque potentiellement important lié à la consommation de la denrée, une option consisterait à maintenir l'autorisation de mise sur le marché existante, et s'il n'y a plus de risque important pour la santé, une option consisterait à révoquer l'autorisation de mise sur le marché. Dans certains cas, il peut être possible d'étendre l'application des LM pour un groupe de denrées alimentaires, dont une denrée a été exclue, à la denrée exclue si, pour cette denrée, seul un petit nombre de données est disponible. (par exemple, suppression d'une exclusion pour le *Brassica* en conserve d'une LM pour les légumes en conserve, s'il n'y a pas suffisamment d'échantillons pour maintenir la LM pour le *Brassica* en conserve).

Dans le cas où les données disponibles sur les denrées alimentaires individuelles sont insuffisantes, mais que les données sur le groupe de denrées alimentaires sont suffisantes.

86. Même lorsque la taille de l'échantillon est suffisante pour l'ensemble d'un groupe alimentaire, si les données sont séparées en fonction des différents aliments de ce groupe, la taille de l'échantillon peut être faible pour les différents aliments. En général, la question de savoir si des LM doivent être établies pour un ou plusieurs groupes ou sous-groupes de denrées alimentaires ou pour une ou plusieurs denrées alimentaires individuelles doit être tranchée au moment de la préparation d'un document de projet, avant d'entamer de nouveaux travaux sur l'élaboration de LM ou, au plus tard, l'élaboration et l'examen d'un document de travail. Si, après l'appel et la collecte de données, il s'avère que les données disponibles sont moins nombreuses 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 un sous-groupe d'aliments par un groupe d'aliments.
87. La question de savoir s'il convient d'établir une LM pour un groupe d'aliments dépend de la similitude des schémas de distribution des aliments individuels au sein du groupe. 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 deux ensembles de données ou plus), peuvent être utilisés pour déterminer si les modèles de distribution des denrées alimentaires du groupe peuvent être considérés comme provenant de la même population, même lorsque le nombre de données est relativement faible (pour les tests statistiques, voir les paragraphes 99 à 110: Traitement de plusieurs ensembles de données). Si le nombre de données est relativement faible, la comparaison des ensembles de données par des diagrammes en boîte à moustache est également utile, à condition que les données censurées à gauche représentent moins de 25 % de l'ensemble de données concerné.
88. Si un certain aliment individuel présente un schéma de distribution différent de celui d'autres aliments individuels dans le groupe alimentaire comparé, il peut être nécessaire d'établir deux LM différentes, l'un pour le groupe alimentaire excluant l'aliment individuel qui présente un schéma de contamination différent, et l'autre pour l'aliment individuel spécifique qui est exclu et qui est également préoccupant. Des approches/décisions similaires peuvent être prises pour le(s) sous-groupe(s) du groupe alimentaire. Si les données relatives à une ou plusieurs denrées alimentaires sont insuffisantes pour atteindre le nombre minimal d'échantillons requis, des appels de données supplémentaires seront lancés pour les denrées alimentaires pour lesquelles il est jugé

nécessaire d'établir des LM. Si la consommation d'une denrée alimentaire individuelle présentant un schéma de distribution différent de celui du groupe de denrées alimentaires ne contribue pas à l'exposition totale au contaminant concerné et peut être négligeable du point de vue de la protection de la santé du consommateur, aucun appel de données supplémentaires n'est requis et son exclusion de l'application du niveau maximal pour le groupe de denrées alimentaires serait une option (par exemple, le niveau maximal pour le plomb dans le sel, qualité comestible, à l'exclusion du sel provenant des marais). En ce qui concerne les groupes d'aliments et leurs sous-groupes, il est possible de se référer aux denrées couvertes par les normes de denrées pertinentes du Codex, à la classification des aliments destinés à l'alimentation humaine et animale (CXA 4-1989) (utilisée également par le CCPR) et à d'autres systèmes de catégorisation des aliments utilisés par le CCFA pour les aliments transformés.

89. Lors de l'élaboration d'une LM pour un groupe alimentaire plus large en raison de la disponibilité limitée de données pour des aliments ou sous-groupes individuels, s'il n'y a que quelques sous-groupes dont les schémas de distribution peuvent être différents de ceux d'autres sous-groupes du même groupe alimentaire, mais pour lesquels il n'y a pas suffisamment de données pour établir une LM distincte, ces aliments ou sous-groupes alimentaires pourraient être exclus de l'application de la LM.

Traitement des ensembles de données comportant une grande proportion de données censurées à gauche (y compris l'utilisation de méthodes de substitution)

Dans certains cas, les résultats de l'analyse d'un contaminant sont obtenus à l'aide de plusieurs méthodes d'analyse et/ou de la même méthode d'analyse, mais avec des sensibilités très différentes. Par conséquent, lorsque des ensembles de données provenant de différentes sources sont combinés, il peut y avoir un large éventail de limites de détection (LD) et de limites de quantification (LQ) pour ce contaminant et cette matrice alimentaire dans l'ensemble de données combiné.

La pertinence des limites de quantification utilisées pour obtenir des ensembles de données de diverses sources doit être examinée au cours du processus de nettoyage des données, et les ensembles de données jugés inappropriés sont exclus avant l'analyse statistique. Pour l'analyse statistique, il convient de n'utiliser que des ensembles de données provenant de méthodes analytiques dont les limites de quantification sont appropriées (voir le chapitre B pour plus de détails).

La NGCTAHA stipule que «*Les limites maximales devraient, de préférence, ne pas être inférieures aux concentrations auxquelles peuvent s'appliquer les méthodes d'analyse mises en place et pratiquées aisément dans les laboratoires de contrôle de l'alimentation humaine et animale, à moins que des considérations de santé publique ne rendent nécessaire une limite maximale inférieure qui suppose une méthode d'analyse plus complexe et plus sensible, basée sur un seuil de détection adéquat plus faible.*».

90. Dans cette rubrique, le terme «ensemble de données» fait référence à un ou plusieurs ensembles de données qui figurent parmi les ensembles de données sélectionnés pour le développement des LM (voir les points 99 à 110). Cette rubrique est particulièrement pertinente lorsque les ensembles de données d'occurrence utilisés pour le développement de LM après le nettoyage des données contiennent encore un taux élevé de données non quantifiées (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.)
91. 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 rapportées.
92. Pour l'analyse statistique d'ensembles de données contenant des données censurées à gauche, des méthodes de substitution sont traditionnellement envisagées. Si l'ensemble de données contient un taux élevé de données censurées à gauche, l'analyse statistique utilisant uniquement des valeurs finies (valeurs quantifiées) n'est pas recommandée, car cette pratique introduit un biais dans les résultats de l'analyse statistique. Une autre méthode consiste à modéliser la distribution en utilisant uniquement les valeurs finies d'un ensemble de données et à estimer les valeurs des centiles élevées en tenant compte du pourcentage de données censurées à gauche. Comme cette méthode n'a pas été utilisée auparavant par le CCCF, elle n'est pas incluse dans les lignes directrices. (Voir également les paragraphes 121 à 130)

Méthodes de substitution

93. 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
- Estimation ponctuelle entre les deux scénarios extrêmes (estimation basse et haute), le scénario de l'estimation moyenne: calcul effectué en attribuant une valeur de $LOQ/2$, racine carrée de la LOQ, ou $(LOD + LOQ)/2$ si la LOD est connue pour les résultats d'analyse inférieurs à la LOQ rapportée.

Si la LOQ n'est pas indiquée et que seule la LOD l'est, il convient d'utiliser la LOD comme alternative, bien qu'il faille examiner attentivement si les résultats analytiques peuvent être utilisés sans indiquer la valeur de la LOQ de la méthode utilisée.

94. 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 les paragraphes Section D7 et EHC 240²⁵). Le choix de scénario d'estimation basse, haute ou moyenne peut affecter la moyenne arithmétique calculée et l'exposition moyenne estimée sur la base de la moyenne arithmétique. Toutefois, pour l'élaboration d'une LM, à moins qu'une grande majorité des points de données ne soient censurés à gauche (c'est-à-dire <LOD ou <LOQ), l'effet des valeurs censurées à gauche sur le calcul des valeurs de centile élevées est négligeable et il y a peu d'impact sur la dérivation des LM proposées, quel que soit le scénario choisi. (voir rubrique D5).
95. 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, tous les éléments d'estimation basse, haute ou moyenne devraient être calculés et présentés. Il est plus important de connaître le modèle de distribution des valeurs finies que le pourcentage de données censurées à gauche lors de l'estimation des valeurs de centile élevées à l'aide d'une distribution modélisée pour établir les LM.
96. Lorsque la dispersion des valeurs finies 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 n'est pas nécessaire, à moins que le contaminant ne soit hautement toxique. Si l'exposition alimentaire estimée dans le cadre du scénario d'estimation supérieure est bien inférieure à la valeur indicative basée sur la santé (HBGV), 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 une LM ne serait pas nécessaire. Il se peut que les HBGV ne soient pas établies pour certains contaminants. Pour ces contaminants, même si toutes les données sont <LOQ, mais qu'il existe un certain risque pour la santé, la ou les LM seront établies à la valeur LOQ pour le moment. Toutefois, si la plupart des données sont <LOQ et qu'il n'y a pas ou peu de problèmes de santé, il n'est pas nécessaire d'établir de LM. Par exemple, un ensemble de données combinées sur le plomb dans les œufs de poule frais contenait 99 % de données censurées à gauche après nettoyage des données et des valeurs finies allant de 0,001 à 0,257 mg/kg. L'impact calculé sur la réduction de l'exposition à des LM hypothétiques était faible, et une LM proposée se situait dans la fourchette des limites de quantification rapportées. Par conséquent, le développement de la LM pour les œufs de poule a été interrompu (réf. CX/CF 22/15/7).
97. Si l'exposition estimée par le scénario d'estimation supérieure est proche ou supérieure à la HBGV ou si la marge d'exposition n'est pas suffisamment élevée pour le profil de toxicité, l'élaboration d'une LM doit être envisagée même si une LM proposée est proche des LOQ rapportées ou se situe dans la fourchette des LOQ rapportées, à condition qu'il existe une (des) méthode(s) d'analyse validée(s) dotée(s) d'une LOQ appropriée. Si nécessaire, il peut être recommandé de demander des données supplémentaires en utilisant des méthodes d'analyse plus sensibles avec des valeurs de limite de quantification plus faibles.
98. Lorsque les valeurs finies présentent une variation importante et atteignent des valeurs significativement élevées, il est conseillé de développer une LM afin d'éliminer les aliments hautement contaminés du marché international. Si le contaminant est hautement toxique ou génotoxique/cancérogène et qu'il se trouve dans des denrées alimentaires consommées en grandes quantités, une LM serait nécessaire 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 contenait 94 % de données censurées à gauche après nettoyage, et la fourchette supérieure des concentrations quantifiées dans cet ensemble de données dépassait 200 µg/kg. Cela indique qu'une LM basée sur des valeurs de centile élevées aurait un impact important sur la réduction de l'exposition alimentaire aux aflatoxines provenant du sorgho en grains. Un projet de LM a été proposé pour les

²⁵ CRITÈRES DE SANTÉ ENVIRONNEMENTALE 240, Principes et méthodes pour l'évaluation des risques des produits chimiques dans l'alimentation (OMS, 2009)

afatoxines dans le sorgho (réf. CX/CF 22/15/9).

Traitement d'ensembles de données multiples - Décision de combiner ou non les ensembles de données, en particulier lorsque les schémas de distribution sont différents, et analyse des ensembles de données combinés et individuels (par année, par région/pays, par année et par région) (pas de partie distincte: les parties pertinentes de cette partie doivent être intégrées dans la partie «Décision sur la couverture géographique des données d'événements fournies» (paragraphe 60-65) et dans la partie «Décision sur la couverture périodique des données d'événements fournies» (paragraphe 66-71);

Les ensembles de données provenant de différentes régions/continents du monde peuvent présenter des schémas de distribution différents pour diverses raisons (par exemple, des conditions climatiques différentes, des conditions de production différentes, y compris le sol/les techniques, les réglementations locales, etc.)

Étant donné que la NGCTAHA stipule que «*Les propositions de limites maximales pour un contaminant dans les produits alimentaires devraient se fonder sur des données provenant de divers pays et sources, y compris des zones et des processus principaux de production,...*», des ensembles de données combinés ont été utilisés par convention pour élaborer des LM qui seront appliquées à l'échelle mondiale, qu'il y ait ou non des différences dans les schémas de distribution dans chaque ensemble de données.

Pour savoir si les schémas de distribution des concentrations de contaminants diffèrent selon la région ou l'année, il est nécessaire de créer des ensembles de données individuels à des fins de comparaison. Si des ensembles de données régionales ou annuelles sont disponibles pour déterminer la couverture géographique ou annuelle, ils peuvent être utilisés pour comparer les schémas de distribution.

La 16^e session du CCCF a conclu qu'à ce stade, l'ensemble de données global combiné doit être utilisé pour le développement des LM et que les ensembles de données individuels par année ou par région sont fournis pour une considération supplémentaire dans le développement des LM.

Mais à ce stade, aucune ligne directrice n'est donnée quant à l'ensemble de données sur lequel le développement des LM doit se fonder ou à la base de données à laquelle il convient d'accorder la priorité pour le développement des LM (ensemble de données combiné, ensemble de données montrant les schémas de contamination les plus élevés pour autant que la denrée ait été produite selon de bonnes pratiques, ensembles de données provenant des principaux pays ou régions producteurs, ensembles de données provenant des pays importateurs reflétant les niveaux d'un contaminant dans une denrée faisant l'objet d'un commerce international, ensemble de données à utiliser à décider au cas par cas).

Ce sujet doit être discuté après la 17^e session du CCCF(REP23/CF16, para 98 (vi) (b))

99. Comme les LM du Codex sont destinées à une application mondiale, elles devraient idéalement être basées sur des ensembles de données mondiaux. Le CCCF doit décider si la LM doit être basée sur un ensemble de données mondiales ou sur un ensemble de données d'une région ou d'une année spécifique.
100. Les méthodes statistiques recommandées pour comparer les schémas de distribution des ensembles de données individuels par région/pays ou par année comprennent des tests non paramétriques. 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).
101. De nombreux modèles de méthodes statistiques non paramétriques sont disponibles sur Internet. Parmi eux, des modèles MS Excel pour effectuer le test U de Mann-Whitney et le test H de Kruskal-Wallis peuvent être téléchargés à partir du site web de la JMPR de la FAO²⁶.
102. 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 modèles de distribution 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 paragraphes 72-80). Pour un ensemble de données comportant 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 les paragraphes 121-130 pour la méthode de dessin).

Cas où les ensembles de données peuvent être combinés

103. La proposition d'une ou de plusieurs LM à l'aide d'un ou de plusieurs ensembles de données combinés a été faite

²⁶ «Appendix XIV Electronic Attachments (2020_Nov)» et ouvrez «XIV 12 Spreadsheet for Kruskal_Wallis 20 group.xls» pour effectuer le test U de Mann-Whitney et le test H de Kruskal-Wallis.

<https://www.fao.org/agriculture/crops/thematic-sitemap/theme/pests/jmpr/jmpr-docs/fr/>

de manière conventionnelle dans les GTE, sans considérations statistiques. Lorsque les tests statistiques montrent que plusieurs ensembles de données provenant de différentes sources peuvent être issus de la même population, ils peuvent être combinés en vue d'une analyse statistique pour obtenir une LM. Toutefois, des ensembles de données individuels devraient être conservés pour une évaluation plus approfondie de l'impact de la LM (voir points 131-157).

104. Dans certains cas, les tests statistiques montrent que plusieurs ensembles de données ne proviennent pas de la même population, mais que les différences de distribution peuvent être minimales d'après l'inspection visuelle des histogrammes ou des diagrammes en boîte à moustaches. Plus le nombre de points de données inclus dans les ensembles de données est important, plus la puissance statistique est élevée²⁷. Si la différence dans la distribution est faible, par exemple égale à l'incertitude de mesure des méthodes analytiques, chaque ensemble de données peut être combiné et utilisé pour l'analyse statistique.
105. Lorsque le nombre de points de données est significativement différent entre des ensembles de données individuels provenant de régions/pays différents, mais que le test statistique indique qu'ils proviennent de la même population, l'ensemble de données combiné qui en résulte reflète principalement les conditions d'un pays/région dont les points de données sont significativement plus nombreux, plutôt que les conditions des pays/régions qui fournissent les données de manière égale. Pour résoudre ce problème, il serait efficace, bien que nécessitant un processus assez complexe, d'équilibrer les ensembles de données en les pondérant par le ratio de production ou de volume des échanges ou par tout autre facteur raisonnable. La méthodologie et la justification de l'utilisation de la pondération des données doivent encore être examinées.

Cas où des ensembles de données individuels sont utilisés

106. Si l'analyse statistique indique que les modèles de distribution des principaux ensembles de données ne proviennent pas de la même population, les ensembles de données doivent être conservés séparément aux fins de l'analyse statistique pour le développement des LM. Toutefois, cette décision doit être prise au cas par cas, car les différents schémas de distribution dépendent généralement de la denrée spécifique examinée. Il convient de justifier le maintien de l'ensemble de données distinct et, en l'absence de justification, l'ensemble de données combiné peut être utilisé.
107. Lorsque l'on envisage d'utiliser des ensembles de données individuels qui ont été conservés séparément, il est recommandé de comparer les résultats statistiques, tels que les valeurs de centile élevées des ensembles de données séparés à ceux qui ont été combinés, par le biais d'une analyse de sensibilité. Il convient de noter qu'il n'est pas possible d'obtenir des valeurs robustes de centile élevées pour les ensembles de données individuels dont la taille de l'échantillon est inférieure à l'une des trois options précédemment identifiées pour le calcul du nombre minimum d'échantillons requis (voir les paragraphes 72-80).
108. Il convient de noter que lorsque plusieurs ensembles de données sont examinés individuellement, plusieurs LM candidates peuvent être identifiées. Bien qu'il n'entre pas dans le champ d'application des présentes lignes directrices de déterminer quelle valeur candidate doit être sélectionnée en tant que LM, il convient de reconnaître la possibilité d'avoir à traiter plusieurs LM candidates, ainsi que l'application de plusieurs ML.
109. En référence au paragraphe précédent, si les ensembles de données provenant de différentes régions/pays sont analysés séparément à l'aide des méthodes statistiques recommandées dans le présent guide, il est nécessaire de prendre en compte les résultats provenant: des principales régions/pays producteurs et/ou des pays importateurs; de la mise en œuvre des codes de pratique connexes pour la réduction et la prévention des contaminants en question; et/ou de la présence de limites réglementaires pour les contaminants dans les denrées en question. Le GTE chargé de développer les LM devrait discuter des ensembles de données à considérer comme l'ensemble de données principal. Si l'on a l'assurance que les ensembles de données présentant des concentrations élevées concernent des produits fabriqués selon de bonnes pratiques (COP du Codex ou BPA, BPF, etc.), il convient alors de se concentrer sur les ensembles de données présentant des concentrations élevées afin d'envisager des LM applicables à l'échelle mondiale.
110. Que l'analyse des données au sein du GTE chargé d'élaborer les LM utilise un ensemble de données combiné ou des ensembles de données individuels, c'est l'occasion et la responsabilité pour ces membres du Codex de vérifier l'impact du projet de LM (ou des LM hypothétiques) par rapport à leurs propres données (pays) et de fournir des commentaires sur le résultat de leur analyse statistique au GTE.

²⁷ En statistiques, 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)

Détermination et traitement des valeurs aberrantes/extrêmes (pas de partie distincte - les parties pertinentes doivent être intégrées dans les parties pertinentes de la sélection et du nettoyage des données (paragraphes 48-53) et dans la nouvelle partie «valeurs aberrantes/extrêmes» de la rubrique «Sélection et nettoyage des données»)

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.

111. La présence de valeurs aberrantes dans les ensembles de données a un impact significatif sur la moyenne arithmétique et les valeurs maximales, mais pas sur la médiane ni, dans une certaine mesure, sur les valeurs de centile élevées s'il y a un nombre suffisant de points de données fiables. Toutefois, il convient de tenir compte du pourcentage que les valeurs aberrantes potentielles représentent par rapport à l'ensemble des données disponibles. Étant donné que ce sont les valeurs de centile élevées, et non les valeurs maximales, qui sont utilisées comme base pour les LM, l'impact du traitement des valeurs aberrantes sur les LM dérivées est généralement faible, mais dans les cas où un pourcentage notable de points de données (par exemple, 2 à 5 %) est exclu, cela pourrait affecter les interprétations de la possibilité d'obtenir les LM considérées.
112. Les valeurs extrêmes peuvent avoir de nombreuses causes, notamment: des erreurs de mesure et de traitement des données (y compris des calculs incorrects), une erreur humaine dans la déclaration (unité de mesure), un comportement frauduleux (adultération), une variation naturelle du contaminant mesuré (changement

climatique, conditions météorologiques, état du sol, etc.) ou des différences dans les méthodes d'échantillonnage (en particulier pour les mycotoxines à distribution hétérogène).

Avant la détermination des valeurs aberrantes

113. Si, au cours du processus de nettoyage des données, il est déterminé, après consultation de la personne qui soumet les données, que certaines des valeurs extrêmes sont dues à des erreurs de mesure et/ou de déclaration, etc., ces données devront être corrigées ou exclues de l'ensemble des données avant de procéder à l'analyse statistique. Toutefois, les valeurs extrêmes dont la cause n'est pas clairement établie doivent être conservées en tant que valeurs aberrantes possibles dans l'ensemble de données après le nettoyage des données.
114. Étant donné que les valeurs extrêmes peuvent avoir de nombreuses causes et que certaines de ces valeurs dans certains ensembles de données peuvent ne pas être considérées comme des valeurs extrêmes si elles sont combinées avec des données provenant d'autres sources (pays/régions, années différentes, etc.), la question de savoir si une valeur extrême est une valeur aberrante pouvant être exclue doit être évaluée sur l'ensemble de données combiné après le nettoyage. Si l'on décide d'analyser des ensembles de données individuels, il convient d'accorder une plus grande attention à l'exclusion des valeurs extrêmes en tant que valeurs aberrantes.

Test statistique des valeurs aberrantes

115. Certaines approches statistiques permettent d'identifier les valeurs aberrantes par une approche non paramétrique, telle que l'approche de l'intervalle interquartile (IQR). 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.
116. 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 de la taille de l'échantillon 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

117. Une autre façon (arbitraire) 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 $\mu\text{g}/\text{kg}$. 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.

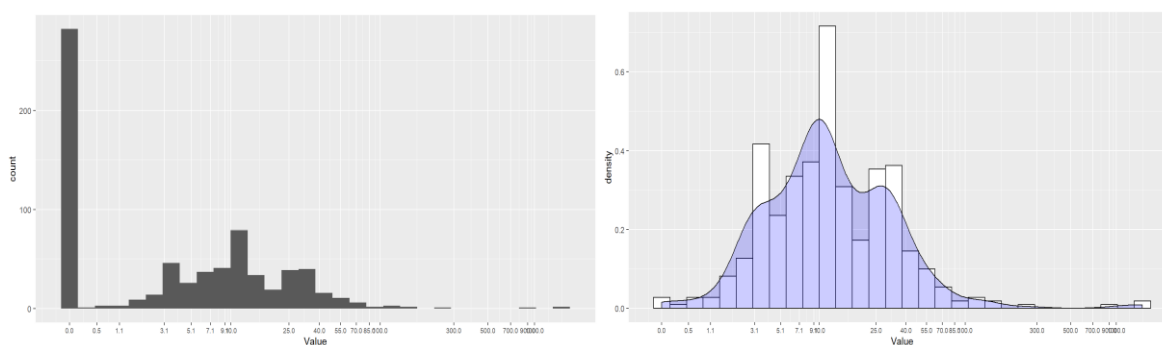


Figure 2. Exemple de données de l'UE sur la somme des toxines T-2 et HT-2 dans les produits de mouture de l'avoine (717 résultats dont 438 résultats quantifiés), (à gauche) Histogramme: tous les résultats, (à droite) Histogramme/courbe de densité de probabilité des résultats quantifiés

Décision sur le traitement d'éventuelles valeurs aberrantes

118. Dans certains cas, les valeurs extrêmes sont scientifiquement valables en fonction des conditions de production, des conditions 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.
119. Les lignes directrices 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, les présentes lignes directrices recommandent que «les valeurs statistiques aberrantes ne soient pas écartées à moins qu'une bonne raison de les exclure ne soit identifiée et expliquée scientifiquement».
120. Néanmoins, si les valeurs extrêmes doivent être exclues en tant que valeurs aberrantes, il est recommandé que la raison de l'exclusion soit clairement indiquée et qu'une analyse de sensibilité soit utilisée 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 qu'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 des valeurs de centile élevées, quelques valeurs extrêmes restant dans l'ensemble de données n'auront que peu d'effet sur le calcul des valeurs de centile.

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*
 - *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 en principe*
 - *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 que l'analyse de sensibilité n'ait un impact important et significatif sur les statistiques récapitulatives. En cas de différence significative dans les résultats de l'analyse de sensibilité, le GTE ou le CCCF décidera au cas par cas si elle 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 les données extrêmes qui nécessitent un examen plus approfondi, le cas échéant. Les statisticiens doivent recommander des tests de valeurs aberrantes adaptés à l'ensemble de données pour l'utilisation par le CCCF, si nécessaire. L'approche IQR peut être une option pour le test des valeurs aberrantes, mais l'exclusion automatique des valeurs aberrantes devrait être*

découragée à moins qu'il n'y ait d'autres justifications pour les exclure. En général, les tests de valeurs aberrantes qui nécessitent une distribution normale ne sont pas recommandés.

Réalisation d'une analyse statistique

L'analyse statistique de base et la présentation des données d'occurrence ont été une pratique courante pour les GTE chargés d'élaborer les LM, mais les rapports sur les résultats des analyses statistiques ont été quelque peu arbitraires en termes de contenu et de format.

Ces lignes directrices présentent les méthodes permettant de dessiner des graphiques et des diagrammes pour montrer les schémas de distribution des données d'occurrence et les valeurs statistiques nécessaires à l'examen et à la discussion de l'adéquation de la LM proposée.

Les analyses statistiques présentées ici sont des exemples et ne sont ni exhaustives ni obligatoires. En fonction du type de contaminant et du nombre de points de données disponibles, des analyses statistiques peuvent être réalisées au cas par cas.

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

121. 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. 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.
122. 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.

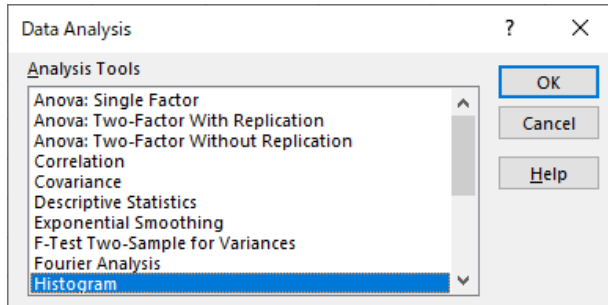


Figure 3. Exemple de menu de l'outil d'analyse des données dans Microsoft Excel

123. 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».
124. En général, les histogrammes donnent une bonne indication des schémas de distribution lorsque les données sont suffisamment nombreuses (environ 50 ou plus). Le nombre approximatif de données nécessaires pour dessiner un histogramme peut servir de guide pour le nombre minimum de points de données nécessaires pour calculer les valeurs de centile élevées (voir les paragraphes 72 à 80). Si le nombre de points de données contenus dans l'ensemble de données diffère, l'axe vertical doit représenter la fréquence relative pour faciliter la comparaison.

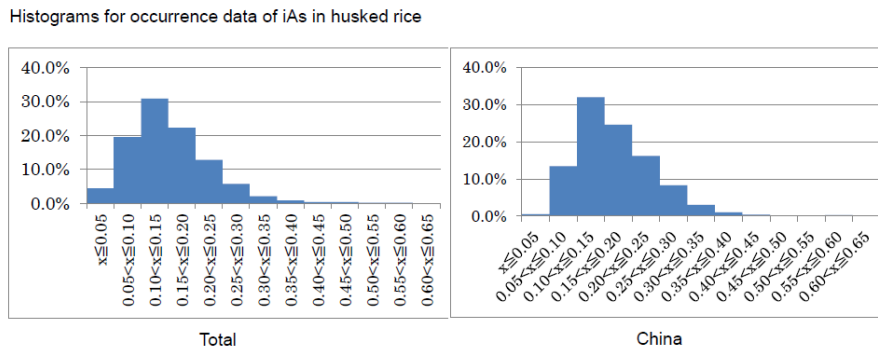


Figure 4. Exemple d'histogrammes de données sur la présence d'arsenic inorganique dans le riz décortiqué (ensemble de données combinées et individuelles) (réf. CX/CF 15/9/7)

- 125. 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.).
- 126. Pour un ensemble de données comportant un 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, car il peut être créé même lorsque le nombre de points de données est de 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 II.1 Box plot for distribution of inorganic arsenic concentration in husked rice in each country

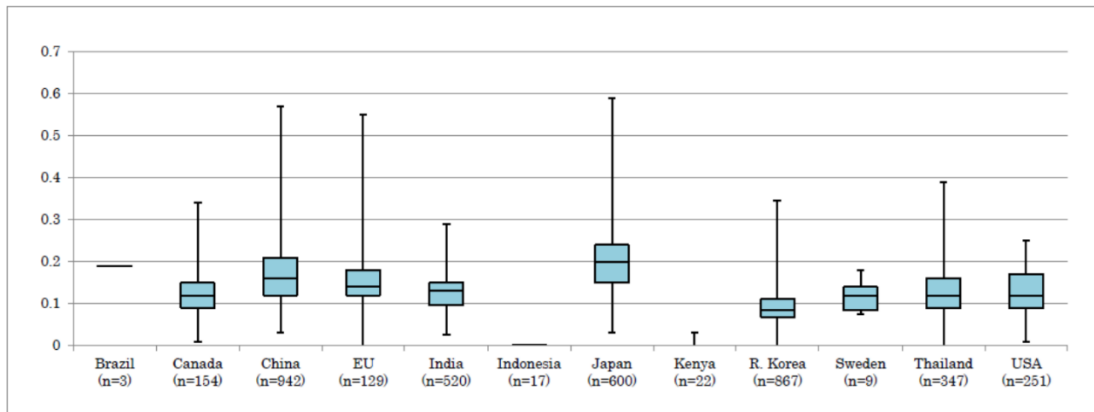


Figure 5. Les diagrammes en boîte à moustaches des ensembles de données individuels sur l'arsenic inorganique dans le riz décortiqué (réf. CX/CF 16/10/5)

- 127. Après avoir tracé des histogrammes ou des diagrammes en boîte à moustaches, il est nécessaire de vérifier s'il existe des valeurs aberrantes et des différences dans: les modèles de distribution des ensembles de données individuels, les formes de la distribution, 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 combiné présente des valeurs aberrantes éventuelles ou une distribution multimodale évidente, il est nécessaire de revenir au processus précédent (tel que la rubrique D3 ou D4) pour reconsidérer la manière de traiter l'ensemble de données.

Agrégation des données et calcul des statistiques descriptives

- 128. Les informations et les statistiques récapitulatives suivantes peuvent être présentées comme un résumé d'un grand nombre de données relatives à l'événement:
 - Nombre de points de données totaux;
 - Nombre de points de données inférieurs aux LOQ rapportées, et/ou rapport entre le nombre de données <LOQ sur le nombre total de points de données;
 - Plage des LOQ rapportées (pour les LOQ appropriées, voir le chapitre «Sélection des données»);

- Moyenne (moyenne arithmétique): si l'ensemble de données contient des points de données inférieurs à la LOQ, trois moyennes arithmétiques basées sur trois scénarios de substitution LB, MB et UB peuvent être préparées (si la distribution est normale ou proche de la normale et symétrique);
 - Si la distribution est très asymétrique, la moyenne géométrique est calculée en utilisant la même approche que ci-dessus;
 - 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, 95e, 97e et 98e 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 échantillons sont inférieurs à la LOQ, les centiles associés pourraient être indiqués comme «<LOQ» (ou LOQ);
 - Minimum;
 - Maximum; dans les cas où le maximum a été identifié comme une valeur aberrante potentielle et que la valeur maximale n'a pas encore 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 (valeurs standard non biaisées), qui est une mesure de l'ampleur de la variation et qui est utilisé comme paramètre pour les fonctions de probabilité telles que les distributions normales, log-normales et gamma; et
 - Les valeurs interquartiles, qui sont une mesure de la variation d'une distribution non paramétrique.
129. 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. 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 de l'application utilisée pour le calcul.
130. 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 LB, MB ou UB. Bien que ce soit au cas par cas, en fonction du nombre de valeurs finies disponibles et du modèle de distribution, 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. (Une application de simulation de modèle est également nécessaire et les détails de son utilisation ne sont pas décrits ici)

Calcul des taux de rejet à des LM hypothétiques

L'identification des taux de rejet appropriés lors de l'établissement des LM n'entre pas dans le champ d'application des présentes lignes directrices.

Le calcul des taux de rejet est une question distincte de la sélection d'un taux de rejet approprié, il constitue l'un des processus nécessaires au développement des LM, il est donc décrit dans cette rubrique.

La présentation des taux de rejet calculés est une pratique courante pour les GTE chargés de développer les LM dans le CCCF, mais dans certains cas, les méthodes et procédures de calcul n'ont pas été clairement énoncées dans le rapport.

Estimation des LM hypothétiques

131. À partir d'une valeur de centile élevée (généralement la valeur du 95e centile) de l'ensemble de données cible pour l'analyse des données, on identifie une valeur candidate pour une LM, en tenant également compte de la précision de la méthode d'analyse actuelle et des chiffres significatifs des résultats d'analyse (par ex, lorsqu'une valeur de centile calculée est de 0,485 mg/kg, la valeur utilisée comme LM candidate (LM hypothétique) sera arrondie à un seul chiffre significatif, par exemple 0,5 mg/kg, et si une méthode analytique de haute précision est disponible ou si la concentration est supérieure d'un ordre de grandeur ou plus, une valeur sera arrondie à deux chiffres significatifs, par exemple 1,0 mg/kg). Il y a quelques exceptions à cette règle où une valeur médiane des valeurs habituellement utilisées pourrait être préférée, comme 0,15 mg/kg, qui, bien que se terminant par

le chiffre «5», pourrait occasionnellement être choisie lorsque des valeurs de 0,10 et 0,20 seraient moins appropriées (par exemple, 0,35 mg/kg pour l'arsenic inorganique dans le riz décortiqué, 0,15 mg/kg pour le plomb dans le vin muté).

132. Une fois que la valeur numérique candidate d'une LM a été déterminée, les valeurs élevées ou faibles les plus proches sont également utilisées comme LM hypothétiques (pour l'exemple ci-dessus, les LM hypothétiques supplémentaires seraient 0,4 et 0,6 mg/kg). En cas de révision de LM existantes, celles-ci doivent également être ajoutés à la liste des LM hypothétiques. En outre, les valeurs obtenues en arrondissant les 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.
133. 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 différentes.
134. Il n'y a pas de règle concernant le nombre de LM hypothétiques à proposer, mais il est préférable d'identifier au moins 2 à 4 valeurs, en fonction de la condition, afin d'examiner leurs effets sur la réduction de l'exposition alimentaire et l'impact économique découlant des taux de rejet, qui seront discutés plus en détail au sein du GTE chargé d'élaborer les LM et lors de la réunion ultérieure du Comité.

Calcul des taux de rejet à des LM hypothétiques

135. 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 rapport.

$$\text{Taux de rejet (\%)} = (\text{nombre d'échantillons} > \text{LM hypothétique}) / (\text{nombre total d'échantillons}) \times 100$$

136. Il convient de noter que le taux de rejet obtenu peut être différent de celui prévu pour le centile supérieur en raison du processus d'arrondissement. Plus le nombre d'échantillons dans l'ensemble 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

137. 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 de calculer les taux de rejet pays par pays, car cela risque de mettre en évidence, à tort, certains aspects économiques non liés à la base scientifique du développement des LM et d'accroître la charge de travail des analystes de données.
138. Pour les contaminants dont on sait que les concentrations varient fortement d'une année à l'autre, le taux de rejet sera 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

139. Pour différentes LM hypothétiques, les taux de rejet calculés peuvent ne pas changer ou changer de manière significative, en fonction du modèle de distribution. La fréquence des 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).
140. 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 (par exemple, arsenic inorganique dans le riz, réf. CX/CF 16/10/5, CX/CF 15/9/7, CX/CF 14/8/6), et des estimations plus fines et améliorées du taux de rejet peuvent être possibles.
141. Une telle approche nécessite davantage de données et de ressources et peut être menée si la charge de travail des GTE le permet. Si une évaluation d'impact plus détaillée ou améliorée concernant les taux de rejet est nécessaire, il est possible de demander une évaluation au Comité mixte FAO/OMS d'experts des additifs

alimentaires (JECFA) (par exemple, les AFT dans les arachides prêtes à consommer, le Cd dans les produits à base de cacao).

Calcul des effets des LM sur la réduction de l'exposition alimentaire à des LM hypothétiques

La NGCTAHA stipule que «*pour promouvoir l'acceptation des LM du Codex, il est donc important que les évaluations de l'impact de ces LM sur l'exposition alimentaire soient effectuées de manière cohérente et réaliste. La procédure comprend l'évaluation de l'apport alimentaire par rapport aux LM proposées ou existantes et à la valeur toxicologique de référence. Les propositions de LM doivent être accompagnées de calculs d'apport et de conclusions d'évaluation des risques concernant leur impact sur l'apport alimentaire et l'utilisation...*».

Cela indique clairement que des orientations sur l'évaluation de l'impact des LM proposés sur l'exposition alimentaire sont nécessaires pour le CCCF.

Le calcul et la présentation des effets des LM sur la réduction de l'exposition alimentaire ont été une pratique courante pour les GTE chargés d'élaborer des LM au sein du CCCF récemment, mais dans certains cas, les méthodes et procédures de calcul n'ont pas été expliquées dans les rapports.

Un examen plus approfondi du rôle du Comité dans le calcul des taux de réduction de l'exposition alimentaire lors de l'examen des LM est nécessaire. (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)).

Calcul de l'exposition alimentaire et de son taux de réduction à des LM hypothétiques

142. Pour s'assurer que la LM proposée est appropriée pour la protection de la santé des consommateurs, il est nécessaire d'évaluer quantitativement l'effet d'une LM hypothétique sur la réduction de l'exposition alimentaire au produit cible en comparant l'exposition avec et sans LM (pour les LM hypothétiques, voir la rubrique D6). Dans le cas d'une révision d'une LM existante, l'exposition à la LM déjà établie est comparée à l'exposition aux nouvelles LM hypothétiques (LM révisées).
143. Pour tous les contaminants, l'exposition alimentaire à long terme/chronique peut être calculée à l'aide de l'équation suivante:

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

$$\text{Exposition alimentaire } (\mu\text{g/kg-mc/jour}) = \text{concentration } (\mu\text{g/g}) \times \text{consommation alimentaire (g/kg-mc/jour)}$$
144. 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.
145. 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.
146. 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.
147. 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. 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.

148. 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 URL²⁸ du programme GEMS/Aliments. 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.
149. 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 CIFOCCO, etc.). Les données utilisées et la manière dont elles ont été traitées devraient être indiquées dans le rapport du GTE.
150. 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.

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

151. 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 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.
152. 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.
- $$\text{Rapport à la VABS (\%)} = \text{exposition à la LM } (\mu\text{g/personne/jour}) / \text{poids corporel moyen (kg/personne)} / \text{VABS } (\mu\text{g/kg pc/jour}) \times 100$$
- Si les VABS sont fixées par semaine ou par mois, l'exposition doit être multipliée par le facteur approprié, par exemple 7 ou 30.

Évaluation de l'impact d'une LM sur l'exposition alimentaire

153. Tous les GTE chargés d'élaborer des LM devraient évaluer l'équilibre entre le taux de rejet et le taux de réduction de l'exposition alimentaire pour chaque LM hypothétique, et déterminer la limite la plus basse qu'il est raisonnablement possible d'atteindre ou proposer des options au Comité pour éclairer cette décision.
154. Bien qu'il n'entre pas dans le champ d'application de ce guide de déterminer quel taux de rejet est le plus approprié, le GTE chargé d'élaborer des LM 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.

Amélioration du calcul des taux de réduction de l'exposition

155. Pour différentes LM hypothétiques, les taux de rejet calculés peuvent ne pas changer ou changer de manière significative, en fonction du modèle de distribution. La fréquence des données dans la plage de centiles élevés est souvent plus faible que dans celle des centiles faibles, ce qui affecte l'estimation des LM hypothétiques et des taux de réduction de l'exposition (voir la Figure 1 pour la forme de la distribution).
156. Si le schéma de distribution de l'ensemble de données combiné (potentiellement mondial) présente un pic unique, des applications de modélisation/simulation (telles que @Risk, Crystal Ball, R, etc.) peuvent être utilisées pour modéliser la distribution afin d'estimer en continu la distribution près des valeurs de centile élevées de la

²⁸ <https://www.who.int/teams/nutrition-and-food-safety/databases/global-environment-monitoring-system-food-contamination>

fonction de distribution, et de produire facilement une concentration moyenne sous une LM hypothétique en appliquant une valeur limite arbitraire à la courbe de distribution modélisée.

157. Une telle approche nécessite davantage de ressources, et peut être menée si la charge de travail des GTE le permet. Si une évaluation d'impact plus détaillée ou améliorée concernant l'exposition alimentaire est nécessaire, il est possible de demander une évaluation au JECFA (par exemple, aflatoxines totales dans les arachides prêtes à consommer, Cd dans les produits à base de cacao).

PRESENTATION DES DONNEES DANS LES RAPPORTS DES GTE AU CCCF

158. Il est important que les données soient présentées dans le rapport du GTE au CCCF de manière à permettre une discussion éclairée sur les LM appropriés à délibérer dans le cadre de la procédure par étapes. 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 sont gérées, si les données en dehors de la base de données GEMS/Aliments ont été prises en compte, etc. Le rapport devra être accompagné d'une justification détaillée.

Présentation de l'analyse des données: analyse statistique

159. Cette rubrique 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 le développement de LM. Le GTE peut utiliser les modèles ou les modifier au cas par cas, car le détail du rapport dépend de la quantité de données disponibles et de la nature du contaminant.
160. 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)
161. Le Tableau 2 peut servir de modèle pour la présentation des éléments ci-dessus.

Tableau 2. Exemple de modèle: Résumé des statistiques de base des ensembles de données

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

Notes au Tableau 2: À remplir pour l'ensemble de données combiné (potentiellement l'ensemble de données global), centile: valeur de centile.

162. Si nécessaire, l'inclusion des éléments suivants est également utile pour l'interprétation de l'ensemble de données:
- plage des LOQ rapportées; et
 - l'écart-type (non biaisé).
163. 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 et, en cas de différence statistiquement significative dans le modèle de distribution, l'analyse devrait envisager de présenter les données par région géographique (par exemple, continent ou région du Codex) à l'aide du Tableau 3.

Tableau 3. Exemple de modèle: Résumé des statistiques de base des ensembles de données par région ou par année

Région/ pays/ Monde	Nombre total d'échantil- lons	Nombre de <LOQ	Moyenne (LB-UB) (mg/kg)	Médiane (mg/kg)	95 ^e centile (mg/kg)	97 ^e centile (mg/kg)	98 ^e centile (mg/kg)	Min. (mg/kg)	Max. (mg/kg)

Notes au Tableau 3a: À remplir par région ou au niveau mondial, centile: valeur du centile.

Année	Nombre total d'échantil- lons	Nombre de <LOQ	Moyenne (LB-UB) (mg/kg)	Médiane (mg/kg)	95 ^e centile (mg/kg)	97 ^e centile (mg/kg)	98 ^e centile (mg/kg)	Min. (mg/kg)	Max. (mg/kg)

Note au Tableau 3b: À compléter par année/mondial ou par année/région, centile: valeurs de centile

164. Si le nombre d'échantillons est supérieur au nombre minimum d'échantillons requis pour un aliment donné ou si la distribution des concentrations pour un aliment donné est significativement différente de celle des autres aliments malgré le nombre inférieur d'échantillons, 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.
165. En général, 2 à 4 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 4.
166. Si nécessaire, un tableau, utilisant le Tableau 4 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 4. 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 d'échantillons)				

Notes au Tableau 4: Réduction de l'exposition = pourcentage de diminution de l'apport alimentaire du contaminant provenant de l'aliment ou du groupe d'aliments concerné

167. Enfin, un document de travail préparé pour être discuté lors de la session plénière du CCCF devrait être accompagné d'un tableau, préparé en utilisant le Tableau 5 comme modèle et conformément au format de la

NGCTAHA, qui comprend des informations sur l'avant-projet/le projet proposé de LM ainsi que des notes explicatives pour les produits de base examinés.

Tableau 5. Format du tableau dans la NGCTAHA

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

APPENDICE

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 limitée	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 denrées alimentaires de base individuelles 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	<p>La distribution de probabilité, $f(x)$, donnée par</p> $f(x) = \frac{x^{\gamma-1} \exp(-x/\beta)}{\beta \Gamma(\gamma)}, \quad 0 \leq x < \infty, \beta > 0, \gamma > 0$ <p>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.</p> <p>mean = $\beta\gamma$ variance = $\beta^2\gamma$ skewness = $2\gamma^{-1/2}$ kurtosis = $3 + \frac{6}{\gamma}$</p> <p>La distribution de $u = x/\beta$ est la distribution gamma standard avec la fonction de densité correspondante donnée par</p> $f(u) = \frac{u^{\gamma-1} e^{-u}}{\Gamma(\gamma)}$ <p>La fonction Γ définie par</p> $\Gamma(r) = \int_0^{\infty} t^{r-1} e^{-t} dt$ <p>Où $r > 0$ (r n'est pas nécessairement un nombre entier). La fonction est réursive et satisfait à la relation.</p> $\Gamma(r+1) = r\Gamma(r)$ <p>(Cambridge Dictionary of Statistics)</p>
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)
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)

Terme	Définition / Explication
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 résidu 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 résidu 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 (JMPPR)	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)
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)

Terme	Définition / Explication
Test H de Kruskal-Wallis	<p>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</p> $H = \frac{12 \sum_{i=1}^k n_i (\bar{R}_i - \bar{R})^2}{N(N-1)}$
Données censurées à gauche	Données qui ne sont pas des valeurs limitées (valeurs quantifiées) ou qui sont inférieures aux LOQ ou LOD déclarées.
Limite de détection (LOD)	<p>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)</p> <p>OU</p> <p>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)</p>
Limite de quantification (LOQ)	<p>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: $LOQ = kQ \sigma_Q$, $kQ = 1/RSD_Q$ 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)</p> <p>OU</p> <p>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)</p>

Terme	Définition / Explication
Distribution log-normale	<p>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</p> $f(x) = \frac{1}{x\sigma(2\pi)^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma^2}(-\ln x - \mu)^2\right] \quad 0 \leq x < \infty$ <p>La moyenne, la variabilité, l'asymétrie et l'aplatissement de la distribution sont</p> $\text{mean} = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$ $\text{variance} = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$ $\text{skewness} = (\exp(\sigma^2) + 2)(\exp(\sigma^2) - 1)^{\frac{3}{2}}$ $\text{kurtosis} = \exp(4\sigma^2) + 2\exp(3\sigma^2) + 3\exp(2\sigma^2) - 3$ <p>Pour les petits σ, la distribution est approximée par la distribution normale. (Cambridge Dictionary of Statistics)</p>
Limite maximale (LM)	<p>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)</p>
Test U de Mann-Whitney	<p>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)</p>
Incertitude de mesure	<p>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, modifiées en 2021)</p>
Moyenne (Arithmétique)	<p>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</p> $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ <p>Très utile lorsque les données ont une distribution symétrique et ne contiennent pas de valeurs aberrantes. (Cambridge Dictionary of Statistics)</p>
Moyenne (Géométrique)	<p>Une mesure de la localisation, g, calculée à partir d'un ensemble d'observations x_1, x_2, \dots, x_n comme</p> $g = \left(\prod_{j=1}^n x_j\right)^{\frac{1}{n}}$ <p>(Cambridge Dictionary of Statistics)</p>
Médiane	<p>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)</p>

Terme	Définition / Explication
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	<p>Une distribution de probabilité, $f(x)$ d'une variable aléatoire, X, qui est supposée par de nombreuses méthodes statistiques. Spécifiquement donné par</p> $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right]$ <p>Où μ et σ^2 sont, respectivement la moyenne et la variabilité de x. (Cambridge Dictionary of Statistics)</p>
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 lignes directrices)
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 associée à des aliments autrement sains et nutritifs. (CXS193-1995)
Assurance qualité (dans les laboratoires d'analyse)	<p>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)</p> <p>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)</p>
Taux de rejet/taux d'infraction	Taux de rejet (ou taux d'infraction) (%) = (nombre d'échantillons présentant des concentrations d'un contaminant supérieures à la LM)/nombre total d'échantillons x 100.

Terme	Définition / Explication
Asymétrie	L'absence de symétrie dans une distribution probabiliste. (Cambridge Dictionary of Statistics)
Écart-type	<p>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:</p> $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ <p>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)</p>
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 V**Liste des participants****PRÉSIDENT****Union européenne**

Frans Verstraete
Commission européenne

COPRÉSIDENTS**Japon**

Tetsuo Urushiyama
Ministry of Agriculture, Forestry and Fisheries

Pays-Bas

Nikki Emmerik
Ministry of Health

Astrid Bulder
RIVM

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 Tecnología 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

BELGIQUE

Christine Vinkx
FPS Health, Food Chain Safety and Environment

Andrea Carletta
FPS Health, Food Chain Safety and Environment

BRÉSIL

Larissa Bertollo Gomes Porto
ANVISA

Giselle Kindlein
Ministry of Agriculture

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
Commission européenne

FRANCE

Jean-Cédric Reninger
ANSES.

Niels Enslin
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

Mansooreh Mazaheri
ISIRI-Standard Research Institute

JAPON

Codexjapan
Ministry of Health, Labour and Welfare

Yoshiyuki Takagishi
Ministry of Agriculture, Forestry and Fisheries

Tomotaro Yoshida
Ministry of Agriculture, Forestry and Fisheries

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

NOUVELLE-ZÉLANDE

Jeane Nicolas
Ministry for Primary Industries, New Zealand Food

Fiapaipai Ruth Auapaau
Ministry for Primary Industries

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

Yeon Ju Kim
Ministry of Food and Drug Safety

FÉDÉRATION DE RUSSIE

Darya Bagreeva
Federal Scientific Center of Hygiene

ARABIE SAOUDITE

Mohammed Mousa Ali Alshehri
Saudi Food and Drug Authority

Lama Almaiman
Saudi Food and Drug Authority

SINGAPOUR

How Chee Ong
Singapore Food Agency

Er Jun Cheng
Singapore Food Agency

Peggy Chew
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 Pinar
Ministry of Agriculture and Forestry

Sinan Arslan
Ministry of Agriculture and Forestry

OUGANDA

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

Quynh-Anh Nguyen
FDA/CFSAN

URUGUAY

Santiago Viera
Laboratorio Tecnológico del Uruguay

Natalie Merlinski
Ministry of Livestock, Agriculture and Fisheries

ORGANISATIONS INTERNATIONALES**INTERNATIONAL COUNCIL OF BEVERAGES ASSOCIATIONS**

Simone SooHoo
Maia Jack

INTERNATIONAL ORGANISATION OF SPICE TRADE ASSOCIATION

Shannen Kelly

INTERNATIONAL SPECIAL DIETARY FOODS INDUSTRIES

Marian Brestovansky
Jean Christophe Kremer

FOODDRINKEUROPE

Alejandro Rodarte

THIE | TEA & HERBAL INFUSIONS EUROPE

Farshad Rostami

CODEX SECRETARIAT

Peter Di Tommaso