

**CONSEILS SUR L'ANALYSE DES DONNÉES POUR
L'ÉLABORATION DE LIMITES MAXIMALES ET DE L'AMÉLIORATION DE LA COLLECTE DES DONNÉES¹**

TABLE DES MATIÈRES	Paragraphes
Préambule	1-7
Collecte et transmission des données	8-16
Compléter le modèle GEMS/Aliments	17-52
Extraction des données	53-57
Sélection de données/nettoyage des données	
Considérations générales	58
Origine des données	59-60
Nettoyage des données	61-65
Traitement de données	
Informations manquantes	66-73
Informations erronées	74-77
Données pour lesquelles l'information sur la partie analysée n'est pas claire	78
Données provenant d'échantillons soupçonnés d'être frauduleux ou économiquement falsifiés	79
Données provenant d'un échantillonnage ciblé	80-81
Détermination et traitement des valeurs aberrantes/extrêmes	82-92
Considérations relatives à la limite de quantification (LOQ) et à la limite de détection (LOD)	
Exclusion basée sur la LOQ	93-96
Somme des composants et LOQ	97
Analyse des données: génération d'un aperçu de données	
Aperçu des pays, nombre de points de données, période de couverture des données	98
Couverture géographique des données d'occurrence fournies	99-104
Période couverte par les données d'occurrence fournies	105-110
Analyse statistique des données d'occurrence et traitement d'ensembles de données pour l'élaboration des LM	
Considérations générales	111-113
Nombre minimum de points de données pour l'estimation des valeurs de percentile élevé	114
Traitement des ensembles de données	
Traitement des ensembles de données avec un faible nombre de points de données	115-119
Traitement des ensembles de données lorsque les données disponibles sur les produits de base/aliments individuels sont insuffisantes, mais que les données de pour le groupe d'aliments sont suffisantes	120-123
Traitement d'ensembles de données (y compris l'utilisation de méthodes de substitution) comportant une grande proportion de points de données censurés à gauche	124-126

¹ Le texte bleu entre crochets simples indique les modifications basées sur les changements qui pourraient être mis en œuvre dans une première phase.
Le texte violet entre doubles crochets indique les modifications qui pourraient être apportées ultérieurement.

Méthodes de substitution pour les ensembles de données comportant une grande proportion de données censurées à gauche	127-130
Décision relative au traitement d'ensembles de données multiples	
Analyse des ensembles de données individuels et combinés, et prise de décision sur la nécessité de comparer les ensembles de données individuels avant de les combiner, en particulier lorsque les schémas de distribution sont significativement différents.	131-136
Cas où l'analyse d'ensembles de données individuels est recommandée	137-140
Réalisation d'une analyse des données par visualisation	141
Agrégation des données et calcul des statistiques descriptives	142-144
Calcul des taux de rejet à des LM hypothétiques	
Estimation des LM hypothétiques	145-148
Calcul des taux de rejet à des LM hypothétiques	149-150
Évaluation de l'impact d'une LM sur le taux de rejet	151-152
Amélioration du calcul des taux de rejet	153-155
Calcul préliminaire des effets des LM sur la réduction de l'exposition alimentaire au contaminant d'un ou de produits de base cibles, à des LM hypothétiques	
Calcul de l'exposition alimentaire et de la réduction à des LM hypothétiques	156
Amélioration du calcul des taux de réduction de l'exposition	157
Préparation des recommandations finales au CCCF	158-164
Présentation des données dans les rapports des GTE au CCCF	
Présentation de l'analyse des données: analyse statistique	165-166
ANNEXES²	

² À développer.

PRÉAMBULE

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

COLLECTE ET SOUMISSION DE DONNÉES

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

9. La première page de la base de données comporte deux onglets, l'un pour la page d'accueil (*Home Page*) et l'autre pour les recherches (*Search*). Pour bénéficier de toutes les fonctionnalités, les membres doivent s'inscrire et se connecter à leur compte. Après s'être connectée, la personne désirant soumettre des données a accès à un onglet de chargement (Upload), qui vient s'ajouter à l'onglet Page d'accueil et à l'onglet Recherche. Elle pourra également accéder à des 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 le site web de GEMS/Aliments ([Nutrition and Food Safety \(who.int\)](http://www.who.int/nutrition)) ou sur les pages GEMS/Aliments associées. Vous trouverez des instructions détaillées dans le document, INSTRUCTIONS FOR ELECTRONIC SUBMISSION OF DATA ON CHEMICALS IN FOOD AND THE DIET (ci-après dénommé «INSTRUCTIONS») sur le site web de GEMS/Aliments. Ce document fournit des instructions concernant l'enregistrement d'un compte, la connexion à la base de données GEMS/Aliments, l'insertion de données dans le 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 l'administrateur de la base de données de GEMS/Aliments. Les questions peuvent porter sur des messages d'erreur lors du chargement, des problèmes d'enregistrement, la manière de nommer les échantillons, les champs obligatoires, la définition des champs, les problèmes de mise en correspondance, etc.
13. Si des questions se posent à propos de la conformité des données par rapport à un appel de données spécifique, la personne désirant soumettre les données doit consulter le Président du GTE et, le cas échéant, le ou les secrétariats du Codex ou du JECFA. Les questions peuvent porter sur le fait de savoir si les échantillons correspondent aux définitions fournies dans l'appel de données ou dans le mandat du GTE.
14. Les personnes qui soumettent des données devraient développer et conserver les métadonnées associées aux soumissions de données. Les métadonnées aideront à répondre aux questions susceptibles d'être posées par le GTE. Les métadonnées peuvent inclure l'année de production, la limite de détection (LOD) et la limite de quantification (LOQ) globales associées à un ensemble de données, des informations figurant sur les étiquettes des produits, des informations sur le lieu de collecte (par exemple, importation ou vente au détail), les noms des collaborateurs qui ont soumis les données et la date à laquelle les données ont été soumises, l'identifiant du lot associé à l'ensemble de données soumis, etc.
15. La qualité des données doit être évaluée par la personne qui les soumet avant qu'elles ne soient introduites dans GEMS/Aliments. Si des questions sérieuses se posent quant à la qualité des données (informations manquantes, analyses suspectes), ne soumettez pas les données tant que vous n'avez pas obtenu de réponses.
16. Si la personne qui soumet les données identifie un problème concernant un ensemble de données après sa soumission, consultez l'administrateur de la base de données 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)³ 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 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

³ 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.

animaux dans l'ensemble de données et à associer la classification nationale des aliments avec la classification de l'OMS et de FoodEx2. L'onglet «2. Mappage de l'aliment» contient l'outil de mappage: l'identifiant local de l'aliment (colonne A, texte libre) et deux niveaux de classification dans des menus déroulants, à savoir niveau 1: Groupe OMS de l'aliment (colonne B) et niveau 2: Identifiant OMS de l'aliment (colonne C). Une fois les champs «Identifiant local de l'aliment», «Groupe OMS de l'aliment» et «Identifiant OMS de l'aliment» remplis, le code OMS de l'aliment, le code FoodEx2 et le nom FoodEx2 sont générés automatiquement dans les colonnes E, F et G de l'onglet 2.

20. Lors de la soumission de données, la fréquence à laquelle chaque aliment doit être mappé dans le 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'Identifiant local de l'aliment (précédemment mis en correspondance avec les codes de l'onglet 2), la concentration chimique, les unités de mesure, la LOD, la LOQ, etc. Étant donné que les Identifiants locaux des aliments ont été mis en correspondance dans l'onglet 2, les colonnes B, C et D de l'onglet 3 seront remplies automatiquement avec les informations issues de la mise en correspondance. La colonne A indiquera automatiquement une erreur si l'un des champs de cet onglet est mal rempli. Les colonnes restantes doivent être remplies en suivant les instructions détaillées fournies dans le document *INSTRUCTIONS*.
23. Veuillez noter que dans le 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. Le texte entre crochets (simples ou doubles) indique les changements qui ont été convenus, mais qui n'ont pas encore été mis en œuvre. Le texte bleu entre crochets simples indique les modifications basées sur les changements qui pourraient être mis en œuvre dans une première phase (le calendrier sera communiqué à la 18^e session du CCCF). Le texte violet entre doubles crochets indique les modifications qui pourraient être apportées ultérieurement.
25. Les paragraphes 26 à 52 se trouvant sous le Tableau explicatif 1 fournissent une orientation supplémentaire sur les champs à l'intention des personnes qui soumettent des données.

Tableau explicatif 1: Champs actuels du modèle GEMS/Aliments

Colonne	Champ	Type de champ/ Menu déroulant	Obligatoire ou Facultatif	Bloc de texte (proposé nouveau ou révisé)
E	Identifiant local de l'aliment	Texte libre	Obligatoire	[Fournir un nom <i>court, mais descriptif</i> de l'aliment, tel que «Hoplostète orange» (par opposition à «Poisson») ou «Riz poli/blanc» (par opposition à «riz».)]
F	N° de série de l'enregistrement	Texte libre	Obligatoire	[Un numéro de série est utilisé pour chaque échantillon. Les données relatives à différents contaminants présents dans le même échantillon doivent porter le même numéro de série. Les institutions nationales devraient coordonner la sélection des numéros de série afin de s'assurer que les numéros sont informatifs et ne font pas double emploi.]
G	[Soumission] Pays/Région/Observateur	Menu déroulant <ul style="list-style-type: none"> (Liste des pays, régions, observateurs) Non spécifié 	[Obligatoire]	[Identifie le pays, la région ou l'observateur (région non spécifiée) qui soumet les données; il ne s'agit pas du pays de production. Si l'observateur ne figure pas dans le menu déroulant, choisissez Non spécifié et notez le nom de l'observateur dans les Remarques]
H	Contaminant	Menu déroulant <ul style="list-style-type: none"> (Liste des contaminants) 	Facultatif	[Veuillez sélectionner un contaminant dans la liste. Un contaminant est requis, mais la saisie manuelle dans «Colonne H: Contaminant» est facultative si un contaminant a été indiqué dans la fiche 1: Début. Si «multiple» est sélectionné dans la fiche 1: Début, la saisie manuelle des contaminants dans le champ H est obligatoire.]
I	Origine de l'aliment	Menu déroulant <ul style="list-style-type: none"> Production intérieure Importation Origine mixte Pas d'informations 	Facultatif	
J	Date d'échantillonnage	Texte libre (JJJ)	Obligatoire	
K	Représentativité de l'échantillon [fiabilité]	Menu déroulant <ul style="list-style-type: none"> Échantillonnage aléatoire [(de routine)] Échantillonnage ciblé Pas d'informations 	Obligatoire	[L'échantillonnage ciblé fait référence au suivi ciblé de découvertes spécifiques de contamination. L'échantillonnage aléatoire (de routine) fait référence à un échantillonnage non ciblé et peut inclure la surveillance de routine ou l'échantillonnage de types d'aliments spécifiques ou de pays importateurs.]
L	Identification en laboratoire	Texte libre	Facultatif	[Le laboratoire qui a effectué l'analyse]

Colonne	Champ	Type de champ/ Menu déroulant	Obligatoire ou Facultatif	Bloc de texte (proposé nouveau ou révisé)
M	Assurance qualité analytique	Menu déroulant <ul style="list-style-type: none"> Assurance qualité interne et normes de référence uniquement. Participation réussie à des essais d'aptitude / [comparaisons interlaboratoires] pertinents au cours de la période d'échantillonnage et d'analyse. Accréditation officielle pour les méthodes pertinentes pendant la période d'échantillonnage et d'analyse. Assurance qualité du laboratoire inconnue. 	Facultatif	
N	Unités de mesure pour les taux de contaminants	Menu déroulant [<ul style="list-style-type: none"> mg/kg (µg/kg) ng/kg pg/kg Bq/kg] 	Obligatoire	[Contrôler soigneusement les unités. S'assurer que les unités choisies dans le menu déroulant correspondent aux résultats de l'échantillon.]
O	LOD	Texte libre	Obligatoire pour les résultats non quantifiés [(c'est-à-dire non détectés)] si la LOQ n'est pas fournie. [[Facultatif]]	Saisir une valeur numérique supérieure à 0 et inférieure à la LOQ. Ce champ contient la limite de détection indiquée par le laboratoire. [La LOD et/ou la LOQ sont obligatoires si le résultat n'est pas fourni, la mention «non détecté» est inscrite dans Résultats (T).]

Colonne	Champ	Type de champ/ Menu déroulant	Obligatoire ou Facultatif	Bloc de texte (proposé nouveau ou révisé)
P	LOQ	Texte libre	Obligatoire pour les résultats non quantifiés si la LOD n'est pas fournie. [[Obligatoire]]	Saisir une valeur numérique supérieure à 0 et à la LOD. Ce champ contient la limite de quantification rapportée par [le] laboratoire. [La LOD et/ou la LOQ sont obligatoires si le résultat n'est pas fourni - la mention «non détectée» est inscrite dans Résultats (T).]
Q	Résultats basés sur	Menu déroulant <ul style="list-style-type: none"> • Teneur en graisses • Poids sec • Tel quel (cru, frais, [, tel que vendu]) • Tel que consommé 	Obligatoire	
R	Portion analysée	Menu déroulant <ul style="list-style-type: none"> • Uniquement comestible • [Aliment entier (comestible + non comestible)] 	Obligatoire	[Exemple: noix décortiquée (comestible) par rapport à noix non décortiquée (aliment entier)]
S	État de l'aliment analysé [[Cuit/Cru]]	Menu déroulant <ul style="list-style-type: none"> • Cuit • Cru • Pas d'informations 	Facultatif	[Exemple: poisson cru par rapport au poisson cuit]
T	Résultats	Texte libre	Obligatoire	[La saisie d'un résultat est obligatoire: soit un zéro, soit non détecté, soit un résultat numérique. Zéro ou non détecté ne peut être indiqué que si la LOQ ou la LOD sont fournies.]
U	Données individuelles par rapport aux données agrégées	Menu déroulant <ul style="list-style-type: none"> • Individuel • Agrégé 	[[Obligatoire]]	
V	Confidentialité des données	Menu déroulant <ul style="list-style-type: none"> • Oui • Vide 	Facultatif	[Toutes les données pour lesquelles «vide» est choisi ou aucune option n'est sélectionnée seront considérées comme non confidentielles dans le cadre du traitement et de l'analyse des données.]
W	Remarques/Références	Texte libre	Facultatif	
X	[[Année de production/récolte]]	[[Texte libre (JJJ)]]	[[Facultatif]]	

Colonne	Champ	Type de champ/ Menu déroulant	Obligatoire ou Facultatif	Bloc de texte (proposé nouveau ou révisé)
Y	[[Informations sur la composition]]	[[Texte libre]]	[[Facultatif]]	[[Informations figurant sur les étiquettes ou déterminées de manière analytique, telles que les principaux ingrédients, la teneur en matières grasses, la teneur en eau ou le pourcentage de cacao total.]]
Z-1	[[Pays/région de production du produit fini]]	[[Menu <ul style="list-style-type: none"> • Pas d'informations • Pays/Régions (A-Z)]]	[[Facultatif]]	
Z-2	Pays/région d'origine des matières premières	[[Menu <ul style="list-style-type: none"> • Pas d'informations • Pays/Régions (A-Z)]]	[[Facultatif]]	
AA	[[Type de produit]]	[[Menu: <ul style="list-style-type: none"> • Destiné à une transformation ultérieure (DFP) • Prêt à consommer (RTE) • Sans objet • Pas d'informations]]	[[Facultatif]]	[[«Destiné à une transformation ultérieure» et «prêt à consommer» sont définis pour certains contaminants et produits dans la <i>Norme générale du Codex pour les contaminants et les toxines présents dans les produits de consommation humaine et animale</i> , CXS-193. Voir paragraphe 48 ci-dessous.]]
BB	[[Lieu d'échantillonnage dans la chaîne de production]]	[[Menu: <ul style="list-style-type: none"> • Pas d'informations • Site de production • Transport de lots en vrac • Frontière (importation/exportation) • Marché/vente au détail • Autres]]	[[Obligatoire]]	[[Fournit des informations sur l'endroit où l'échantillon a été obtenu dans la chaîne de production.]]
CC	[[Principe de la méthode d'analyse]]	[[Menu <ul style="list-style-type: none"> • Méthode A • Méthode B • Méthode Z • Autres • Pas d'informations]]	[[Facultatif]]	

26. **E. Identifiant local de l'aliment.** Champ obligatoire. Dans la mesure du possible, la personne qui soumet les données doit fournir les noms en anglais. Ajouter des détails au nom peut aider l'analyste des données à classer l'échantillon (p. ex., «jus d'ananas-orange» par rapport à «jus»). D'un autre côté, un nom d'échantillon trop long (p. ex. la liste de tous les ingrédients d'un aliment en contenant plusieurs) peut compliquer le travail de l'analyste. Des informations supplémentaires sur le nom peuvent également être ajoutées dans la colonne Remarques.
27. **F. N° de série** Champ obligatoire. Un numéro de série (identifiant de l'échantillon) devrait être utilisé pour chaque échantillon. Si des informations sur plusieurs contaminants sont fournies pour un même échantillon, le même numéro de série doit être utilisé. Les institutions nationales sont chargées de coordonner l'utilisation du même numéro de série pour toutes les soumissions de données d'analyse concernant le même échantillon d'aliment. **Les institutions nationales devraient coordonner la sélection des numéros de série afin de s'assurer que les numéros sont informatifs et ne font pas double emploi.** Les conventions de numérotation doivent fournir des informations qui permettent de remonter à la personne soumettant les données, à la date de soumission ou au programme de surveillance. Éviter de numéroter avec des numéros de série peu informatifs tels que 1, 2, 3, etc. Si les soumissions sont effectuées en réponse à deux appels de données différents, la personne qui soumet le deuxième ensemble de données doit s'assurer que les nouvelles données n'ont pas remplacé les données précédentes, en vérifiant avec l'administrateur de GEMS/Aliments si nécessaire.
28. **G. [Soumission] Pays/Région/Observateur.** [Champ obligatoire. Identifie le pays, la région ou l'observateur (région non spécifiée) qui soumet les données; il ne s'agit pas du pays de production. Si l'observateur ne figure pas dans le menu déroulant, choisissez Non spécifié et notez le nom de l'observateur dans les Remarques]
29. **H. Contaminant.** [Champ facultatif. Un contaminant peut être ajouté sur la «feuille de travail 1»: Début» ou par saisie manuelle dans la «Colonne H: Contaminant. Un contaminant doit être identifié, mais la saisie du nom du contaminant dans la colonne H est facultative si un contaminant a été ajouté sur la fiche 1. Toutefois, si «multiple» est sélectionné dans la fiche 1: Début, la saisie manuelle des contaminants dans le champ H est obligatoire]]
30. **I. Origine de l'aliment.** [Champ facultatif. Identifie les échantillons comme étant d'origine nationale, importée, mixte ou inconnue]
31. **J. Date d'échantillonnage.** [Champ obligatoire. Identifie la date d'échantillonnage.]
32. **K. Représentativité de l'échantillon.** [Champ obligatoire. Le terme «échantillonnage aléatoire (de routine)» fait référence à un échantillonnage non ciblé et peut inclure la surveillance de routine ou l'échantillonnage de types d'aliments spécifiques ou de pays importateurs spécifiques.] Par exemple, les tests effectués sur une large gamme d'échantillons importés d'une certaine catégorie d'aliments pour détecter la présence d'un contaminant donné sont «aléatoires». Le terme «échantillonnage ciblé» devrait être choisi pour les échantillonnages de suivi réalisés à la suite de constatations spécifiques de contamination. Par exemple, si un pays identifie des niveaux élevés d'un contaminant dans un échantillon provenant d'un fabricant particulier, un échantillonnage supplémentaire du même lot ou des lots produits simultanément par le même fabricant serait «ciblé».]
33. **L. [Identification du laboratoire.** Champ facultatif. Identifie le laboratoire qui a effectué l'analyse.]
34. **M. [Assurance qualité analytique.** Champ facultatif. Fournit des informations sur l'assurance qualité analytique du laboratoire d'analyse.]
35. **N. Unités de mesure des taux de contaminants.** [Champ obligatoire. Fournit des unités de mesure pour les résultats des contaminants. Contrôler soigneusement les unités. S'assurer que les unités choisies dans le menu déroulant correspondent aux résultats de l'échantillon. Veiller à ce que l'unité du rapport soit la même pour les résultats, la LOD et la LOQ. Dans le meilleur des cas, la personne qui soumet les données devrait fournir à la fois la LOQ et la LOD, même si ces champs ne sont actuellement obligatoires que pour les résultats non quantifiés.]
36. **O.LOD.** [Le champ LOD contient la limite de détection indiquée par le laboratoire. Saisir une valeur numérique supérieure à 0 et inférieure à la LOQ. Noter que la LOD ou la LOQ sont obligatoires si non détecté a été indiqué dans Résultats (T)] **[[Rapporter la LOD est facultative, mais encouragé.]]** (Commentaire: Les champs O. LOD et P. LOQ doivent être intervertis de manière à ce que le champ P. LOQ obligatoire soit placé en premier).
37. **P. LOQ.** [Le champ LOQ contient la limite de quantification indiquée par le laboratoire. Saisir une valeur numérique supérieure à 0 et à la LOD. Noter que la LOD ou la LOQ sont obligatoires si non détecté est indiqué dans Résultats (T).] **[[La LOQ est obligatoire pour les échantillons soumis après le XX/XX/XXXX. Avant le XX/XX/XXXX, la LOQ n'était obligatoire que pour les résultats non quantifiés si la LOD n'était pas fournie.]]** (Commentaire: Les champs O. LOD et P. LOQ doivent être intervertis de manière à ce que le champ P. LOQ obligatoire soit placé en premier).
38. **Q. Résultats basés sur.** Champ obligatoire. Indique si les résultats sont basés sur une analyse de la denrée «telle quelle» (crue, fraîche, telle que vendue)»; «telle que consommée»; ou basée sur la teneur en matières grasses ou

le poids sec. Exemples: (1) «teneur en matières grasses»: résultats concernant les contaminants liposolubles dans les tissus de viande animale, basés sur la teneur en matières grasses; (2) «poids sec»: résultats concernant les contaminants dans les tissus animaux, lorsque l'échantillon est séché en laboratoire jusqu'à ce que l'humidité soit éliminée avant l'analyse; (3) «tels quels (crus, frais, tels que vendus)»: légumes et fruits secs vendus au détail; également, résultats concernant les préparations en poudre pour nourrissons, non reconstituées avec de l'eau; et (4) «telle que consommée»: résultats concernant les préparations en poudre pour nourrissons diluées.

39. **R. Portion analysée.** Champ obligatoire. Indique si l'aliment entier est analysé ou seulement la partie comestible (par exemple, noix décortiquée (comestible) ou noix non décortiquée (aliment entier)).
40. **S. État de l'aliment analysé.** [Champ facultatif. Indique si l'échantillon alimentaire est cuit ou cru (par exemple, poisson cru ou poisson cuit).]
41. **T. Résultats.** [La saisie d'un résultat est obligatoire: soit un zéro, soit non détecté, soit un résultat numérique. Zéro ou non détecté ne peut être indiqué que si la LOQ ou la LOD sont fournies.]
42. **U. Échantillon global.** Facultatif **[[Obligatoire.]]** Les données agrégées renvoient à des résultats basés sur des échantillons composites, tels que des échantillons provenant d'études de l'alimentation totale. Les données agrégées sont souvent exclues des analyses de taux d'infraction effectuées pour déterminer les LM appropriées, qui sont basées sur l'observation de la distribution des données et des centiles supérieurs dépassant les LM proposées. Toutefois, des données agrégées peuvent être introduites dans la base de données GEMS/Aliments et un nombre restreint de données a été inclus dans les analyses du CCCF par le passé. L'administrateur de la base de données GEMS/Aliments ou un statisticien doit être consulté avant d'inclure des données agrégées. Si des données agrégées sont incluses dans une analyse de LM, cela doit être indiqué dans le document destiné au GTE.
43. **V. Données confidentielles.** [Champ facultatif. **Noter que toutes les données pour lesquelles «vide» a été choisi ou aucune option n'est sélectionnée seront considérées comme non confidentielles dans le traitement et l'analyse des données**] Les pays peuvent soumettre des données «confidentielles» s'ils souhaitent en limiter l'accès à la FAO, à l'OMS et aux organismes techniques apparentés, tels que le Codex. L'administrateur de la base de données GEMS/Aliments peut fournir des enregistrements marqués «confidentiel» aux présidents des GTE; par conséquent, avant de télécharger des données, les présidents des GTE doivent toujours consulter l'administrateur de GEMS/Aliments à propos de l'extraction des données. Si un pays a soumis des données portant la mention «confidentiel» en réponse à un appel de données, il doit également en informer le Président du GTE lors de la phase d'extraction/d'analyse des données.
44. **W. Remarques/Références.** [Champ facultatif qui peut être utilisé pour ajouter des informations pertinentes pour lesquelles il n'existe pas de champ défini. Les informations figurant sur les étiquettes des produits (telles que les ingrédients ou les noms détaillés des produits) ou les informations détaillées sur la méthode d'analyse sont des exemples d'informations qui ont été incluses dans cette colonne. Une référence à un appel de données spécifique est une autre information qui peut être inscrite dans cette colonne.]
45. **[[X. Année de production/récolte.** Champ facultatif. Fournit des informations sur l'année de production, par exemple pour les denrées alimentaires d'origine animale, ou sur l'année de récolte, par exemple pour les denrées alimentaires d'origine végétale, si l'information est disponible.]]
46. **[[Y. Informations sur la composition.** Facultatif. Informations figurant sur **les étiquettes**, telles que les principaux ingrédients, la teneur en matières grasses, la teneur en eau ou le pourcentage de cacao total pour le chocolat.]]
47. **[[Z. Pays/région d'origine/de production.** Facultatif. Fournit des informations sur le pays ou la région d'origine/de production de [l'aliment ou de la matière première ou du lot ou de l'envoi échantillonné]. Si cette information n'est pas connue, il est possible de sélectionner Inconnu.]]
48. **[[AA. Type de produit** Facultatif. Indique si l'échantillon est «destiné à une transformation ultérieure» ou «prêt à consommer», ou cette information est «Sans objet» ou «Inconnue». «Destiné à une transformation ultérieure» ou «prêt à consommer» sont définis à plusieurs endroits dans la [Norme générale du Codex pour les contaminants et les toxines présents dans les produits de consommation humaine et animale](#) (CXS-193-1995). Pour l'aflatoxine dans le maïs, le sorgho, les fruits à coque, les arachides et les figues sèches ou le déoxynivalénol dans les céréales, «destinés à une transformation ultérieure» signifie destinés à subir une transformation/un traitement supplémentaire qui s'est avéré réduire les niveaux d'aflatoxine ou de déoxynivalénol avant d'être utilisés comme ingrédients dans les denrées alimentaires, transformés d'une autre manière ou proposés à la consommation humaine. Pour les aflatoxines dans les fruits à coque et figues sèches, «prêt à consommer» signifie qu'ils ne sont pas destinés à subir une transformation/un traitement supplémentaire qui permettrait de réduire les concentrations d'aflatoxines avant d'être utilisés comme ingrédient dans les produits alimentaires, autrement transformés ou proposés pour la consommation humaine. Pour les produits de base mentionnés ci-dessus, il est

possible de sélectionner «destiné à une transformation ultérieure» et «prêt à consommer» ou «inconnu»; pour les autres produits, il convient de sélectionner «sans objet».

49. La CXS-193-1995 fait également référence à plusieurs endroits à des repas pour nourrissons «prêts à consommer» et à du lait liquide «destiné à une transformation ultérieure». Le champ AA ne s'applique pas à ces utilisations de «prêt à consommer» et «destiné à une transformation ultérieure».]
50. **[[BB. Lieu d'échantillonnage dans la chaîne de production.** Obligatoire. Fournit des informations sur l'endroit où l'échantillon a été obtenu dans la chaîne de production, à savoir: Site de production (par exemple, entreprise, usine alimentaire), transport de lots en vrac, frontière (par exemple, test à l'importation ou à l'exportation), marché/vente au détail, ou autre. Si cette information n'est pas connue, il est possible de sélectionner Inconnu. Si l'option Autre est sélectionnée, des informations peuvent être ajoutées aux Remarques.]]
51. **[CC. Principe de la méthode d'analyse.** Facultatif. La méthode d'analyse, telle que par ICPMS ou par GC-MS, peut être sélectionnée à partir d'un menu déroulant.]]
52. **Erreurs.** Avant d'introduire des données, la personne qui les soumet doit relire le fichier attentivement afin de 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. L'administrateur de la base de données GEMS/Aliments peut être contacté pour obtenir de l'aide.
 - a. Résultat rapporté < LOD, LOQ ou LOD manquante lorsque le résultat est **non détecté** **[-LOQ manquante-]**, LOD > LOQ rapporté
 - b. Dates introduites dans un format incorrect
 - c. Champs obligatoires incomplets
 - d. Doublons dans la fiche en cours de traitement ou dans la base de données

EXTRACTION DE DONNÉES

53. Le processus d'extraction des données commence sur le site web de la base de données: [GEMS/Food Contaminants Database](#). Comme indiqué plus haut, pour bénéficier de toutes les fonctionnalités, les membres doivent s'inscrire et se connecter à leur compte. Après s'être connectés, les analystes verront s'ouvrir une page comportant deux onglets, l'un pour la page d'accueil (*Home Page*) et l'autre pour les recherches (*Search*). L'onglet Page d'accueil contient un petit nombre d'ensembles de données extraits par région et par contaminant. Pour effectuer des recherches spécifiques, l'analyste sélectionne l'onglet Recherche. La fonction de recherche permet à l'analyste de filtrer les données par région de l'OMS, par contaminant, catégorie d'aliments, nom de l'aliment et période d'échantillonnage. Ces filtres lui permettent d'identifier les données répondant à un appel de données ou à du mandat particuliers.
54. Afin d'identifier l'ensemble de données le plus précis à extraire pour le développement des propositions de LM, il est préférable de consulter l'administrateur de la base de données GEMS/Aliments. Lors de la soumission des données, les utilisateurs peuvent faire des choix qui risquent d'entraîner l'omission de données lors de l'extraction. Un exemple: des données téléchargées en tant qu'«aliments pour nourrissons et enfants» peuvent ne pas être prises en compte lors d'une recherche limitée à des «jus de fruits et de légumes». Un autre exemple: des données sur les jus peuvent être associées par erreur aux «fruits et produits à base de fruits», alors que l'identifiant local de l'aliment ou le champ Remarques identifie clairement les échantillons comme étant des jus. La consultation de l'administrateur de base de données de GEMS/Aliments avant l'extraction des données peut aider le GTE à s'assurer qu'il a extrait de GEMS/Aliments toutes les données pertinentes pour l'analyse des LM.
55. Les données confidentielles sont une autre raison pour laquelle, avant de télécharger des données, les présidents des GTE doivent toujours consulter l'administrateur de la base de données GEMS/Aliments. L'administrateur de la base de données de GEMS/Aliments peut fournir les enregistrements portant la mention «confidentiel» aux Présidents des GTE. Ces enregistrements n'apparaîtront pas lors d'une recherche de routine telle que décrite ci-dessus. Les membres des GTE intéressés par une analyse plus détaillée des données confidentielles peuvent consulter le président du GTE.
56. Il est important de conserver un enregistrement de tous les filtres et termes de recherche pour le rapport du GTE.
57. **[[Des modifications ont été apportées à la base de données GEMS/Aliments en 2025/6 afin de rendre la LOQ obligatoire et d'introduire de nouveaux champs tels que X, Y, Z1, Z2, AA, BB, CC. Les données transmises avant ces changements doivent être considérées comme valides et être utilisées pour le développement des LM.]]**

SÉLECTION/NETTOYAGE DE DONNÉES

Considérations générales

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

Origine des données

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

Nettoyage des données

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

Traitement de données

Informations manquantes

66. Une fois que toutes les données non pertinentes ou inappropriées ont été supprimées de l'ensemble de données, les données ne doivent pas être exclues si tous les champs obligatoires sont remplis (voir la section sur la collecte et la soumission des données) et si les données répondent aux critères de téléchargement dans la base de données GEMS/Aliments. Il convient de noter que même si tous les champs obligatoires sont remplis et que les données répondent aux critères, les données peuvent encore être incomplètes pour l'obtention des LM.
67. Dans les cas où l'information sur l'échantillon dans la base de données GEMS/Aliments est nécessaire, mais incomplète (par exemple, manquante ou peu claire), la première étape devrait être de contacter le point de contact du pays/organisation ou de l'observateur qui soumet les données pour permettre d'obtenir l'information manquante. Ceci est particulièrement important si les informations manquantes ont été demandées dans le cadre

d'un appel de données. L'administrateur de la base de données GEMS/Aliments peut également être chargé de cette tâche.

68. Si des informations manquantes sont disponibles, un fichier de données corrigé doit être fourni par le pays/organisation ou l'observateur au GTE et à l'administrateur de la base de données GEMS/Aliments. L'analyse à l'aide des échantillons corrigés pourra se poursuivre. La demande et les corrections doivent être notées dans le document de travail du GTE.
69. Si les informations manquantes ne peuvent être obtenues et que le président du GTE conclut que les données doivent être exclues de l'analyse en raison d'informations manquantes, le GTE doit noter l'exclusion des données et les impacts possibles de l'exclusion dans son document de travail.
70. Exemples d'informations manquantes indiquant que les données devraient éventuellement être exclues d'une analyse plus approfondie:
 - l'état de l'aliment analysé n'est pas précisé (par exemple, cuit par rapport à cru, alors que l'analyse est censée porter uniquement sur des aliments crus)
 - description adéquate du produit dans le champ de l'Identifiant local de l'aliment (par exemple, l'analyse porte sur le «maquereau», mais le produit est décrit comme étant du «poisson», et l'analyse dépend de l'identification de l'espèce de poisson)
 - et autres
71. Exemples d'informations manquantes qui n'empêcheraient pas la poursuite de l'analyse des données (en fonction de l'examen au cas par cas):
 - informations sur l'échantillonnage: type d'échantillonnage, lieu d'échantillonnage dans la chaîne de production
 - l'état du produit, par exemple cru ou cuit, n'est pas indiqué dans le champ «S», mais l'information peut être déduite d'autres informations fournies, par exemple l'échantillon est décrit comme du poisson cuit.
 - principe de la méthode d'analyse utilisée
 - lorsque la LM est basée sur une somme de composants et que les données ne sont pas déclarées pour tous les composants, mais pour ceux qui contribuent de manière significative à la somme ou lorsque l'occurrence est rapportée comme une somme.
72. Le président du GTE ne doit pas exclure toutes les données pour lesquelles il manque un champ facultatif, mais doit examiner les détails nécessaires au développement/à l'élaboration des LM. Par exemple, pour certains produits de base, des informations supplémentaires peuvent être nécessaires pour élaborer les LM. Ainsi, pour les céréales brutes, des informations sur l'étape de transformation peuvent être importantes pour proposer des LM, mais ces informations ne seraient pas nécessaires pour les boissons produites en tant que produits finis.
73. Comme indiqué précédemment, les données non GEMS prises en compte dans le cadre d'une analyse complémentaire peuvent manquer d'informations cruciales. Le président du GTE doit appliquer les mêmes critères et questions (par exemple, l'information manquante est-elle essentielle à l'inclusion) aux données non GEMS.

Informations erronées

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

- 195 résultats dans un ensemble de données de 200 échantillons de denrées alimentaires ayant une teneur en matières grasses typique de 5 % se situent dans la fourchette de 0 à 20 mg/kg; toutefois, 5 points de données se situent dans la fourchette de 100 à 400 mg/kg, ce qui suggère qu'ils ont été rapportés sur la base des matières grasses plutôt que sur la base du poids entier désigné. Lorsque l'on trace ces données dans une courbe de distribution de fréquences, elles sont identifiées comme d'éventuelles valeurs aberrantes (voir paragraphes 82-92 et Annexe I).

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

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

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

79. Pour déterminer si les échantillons contaminés sont le résultat d'une falsification frauduleuse/économique, il faut d'abord tenir compte de la nature du contaminant (par exemple, plomb ou mycotoxine). Des différences importantes d'une année à l'autre peuvent être le résultat de la variabilité naturelle (par exemple, un niveau élevé de mycotoxines dû à des conditions climatiques spécifiques dans une certaine région/année de production). Dans d'autres cas, les écarts importants peuvent être le résultat de mauvaises pratiques (par exemple, le plomb). Les signes possibles d'échantillons frauduleux/économiquement falsifiés sont les suivants:
- certains échantillons sont des ordres de grandeur plus élevés que d'autres, par exemple 0,1 mg/kg par rapport à 100 mg/kg, ou
 - la variabilité temporelle des données (selon le contaminant), par exemple, les données sont beaucoup plus élevées pour une année de l'ensemble de données.

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

Données provenant d'un échantillonnage ciblé et biais

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

Détermination et traitement des valeurs aberrantes/extrêmes

82. Le président du GTE doit examiner l'ensemble des données pour déterminer s'il existe des valeurs aberrantes/extrêmes qui devraient être supprimées. Cette section fournit des orientations sur les différentes approches permettant d'identifier les valeurs aberrantes/extrêmes et les protocoles pour les supprimer, le cas échéant.
83. En règle générale, il est important que les valeurs aberrantes/extrêmes ne soient pas écartées, à moins qu'il n'y ait une raison valable de le faire.
84. Les valeurs extrêmes peuvent être valables, par exemple en raison de l'hétérogénéité de la distribution des contaminants (comme les points chauds pour les mycotoxines) ou en raison de la variation naturelle des contaminants mesurés (par exemple, en raison des conditions météorologiques, de l'état du sol, etc.) D'autres valeurs extrêmes peuvent être erronées, par exemple des erreurs de mesure et de traitement des données, y compris des calculs incorrects ou l'utilisation d'une unité de mesure incorrecte, ou résulter d'un comportement frauduleux (falsification économique).

85. Les valeurs erronées ou les valeurs résultant d'une falsification économique doivent toujours être supprimées de l'ensemble de données final avant de déterminer les LM. Les valeurs extrêmes valides devront être examinées au cas par cas, en utilisant d'abord l'inspection visuelle des données (voir Annexe III pour des exemples), puis les méthodes statistiques (Annexe I).
86. La présence de valeurs aberrantes dans les ensembles de données a un impact significatif sur la moyenne arithmétique et les valeurs extrêmes, mais pas sur la médiane. Toutefois, il convient de tenir compte du pourcentage de valeurs aberrantes potentielles présentes dans l'ensemble des données. Étant donné que les valeurs du centile élevées, plutôt que les valeurs maximales, servent de base à l'élaboration des LM fondées sur les taux de rejet, l'impact des valeurs aberrantes sur les LM dérivées sera généralement faible. Cependant, dans les cas où un pourcentage notable de points de données (par exemple 2-5 %) sont exclus de l'ensemble de données en tant que valeurs aberrantes, cela pourrait affecter les valeurs de centile élevées (voir également le paragraphe 92). Dans ces cas, il convient de fournir un commentaire concernant l'effet de l'exclusion des valeurs aberrantes sur la possibilité d'atteindre les LM (c'est-à-dire le taux de rejet) considérées.
87. L'Annexe I du présent document d'orientation fournit plus de détails sur les approches statistiques permettant d'identifier les valeurs aberrantes.
88. Le terme «valeur aberrante» défini dans la Directive sur la Terminologie analytique (CXG 72-2009) suppose une distribution normale dans un ensemble de données composé de résultats d'analyses répétées du même échantillon. Elle stipule que *«les valeurs statistiques aberrantes sont éliminées à moins que le statisticien ne décide, pour de bonnes raisons, de les conserver»*. En revanche, les ensembles de données traités dans ce document d'orientation sont des résultats d'analyse provenant d'une variété d'échantillons et de différentes méthodes d'analyse. La distribution de ces données étant inconnue (dans de nombreux cas, elle ne suit pas une distribution normale) et les données pouvant être combinées à partir de sources multiples, il est difficile de prévoir l'étendue de la variation au sein d'un ensemble de données. Par conséquent, ce document d'Orientation recommande que les valeurs statistiques aberrantes ne soient pas écartées à moins qu'une bonne raison de les exclure ne soit identifiée et expliquée scientifiquement.
89. Les valeurs extrêmes identifiées comme des erreurs doivent être traitées comme indiqué aux paragraphes 74 à 77. Toutefois, les valeurs extrêmes qui n'ont pas d'explication claire doivent être conservées et évaluées afin de déterminer si elles doivent être traitées comme des valeurs aberrantes dans l'ensemble de données final. Une analyse de sensibilité peut être réalisée pour comprendre l'impact de l'inclusion des valeurs aberrantes sur l'évaluation globale.
90. Étant donné que les valeurs extrêmes peuvent avoir de nombreuses causes et que certaines de ces valeurs peuvent ne pas être considérées comme extrêmes si elles sont combinées avec des données provenant d'autres sources (pays/régions, années différentes, etc.), l'exclusion éventuelle d'une valeur extrême en tant que valeur aberrante doit être déterminée sur la base de l'ensemble de données combinées et nettoyées. Si des ensembles de données individuels sont analysés séparément, il convient d'accorder une plus grande attention à l'exclusion des valeurs extrêmes en tant que valeurs aberrantes.
91. Dans certains cas, les valeurs extrêmes sont scientifiquement valables en fonction des conditions de production et météorologiques, et d'autres facteurs potentiels tels que les éruptions volcaniques, etc. Compte tenu des caractéristiques de la distribution des contaminants des données d'occurrence dans les aliments, il n'est pas recommandé d'exclure simplement les valeurs extrêmes sur la base des résultats des tests statistiques de valeurs aberrantes ou d'autres méthodes telles que l'inspection visuelle. Étant donné que la gamme des concentrations et des schémas de distribution qui peuvent être empiriquement ou théoriquement supposés varie considérablement en fonction du type de contaminant (métaux lourds, mycotoxines, etc.), le traitement des valeurs extrêmes doit être déterminé au cas par cas. Par exemple, une attention particulière doit être accordée aux mycotoxines dont les concentrations peuvent varier de manière significative en fonction des méthodes d'échantillonnage utilisées, en raison de la distribution hétérogène dans un lot, ainsi que des variations annuelles très importantes.
92. Si les valeurs aberrantes sont exclues, il est recommandé que la raison de l'exclusion soit clairement indiquée dans le document final présenté par le GTE à la session plénière du CCCF. Des analyses de sensibilité peuvent être effectuées pour montrer comment l'exclusion ou la non-exclusion des valeurs aberrantes peut ou non affecter le calcul des valeurs de centile élevées. Il convient de rappeler que quelques valeurs extrêmes restant dans l'ensemble de données n'auront que peu d'effet sur le calcul des valeurs de centile élevées, à condition que le nombre total de points de données dans l'ensemble de données soit suffisamment supérieur au nombre minimum de points de données requis pour calculer les valeurs de centile élevées.

Limite de quantification (LOQ) et Limite de détection (LOD)**Exclusion sur la base de la LOQ**

93. Différentes méthodes d'analyse fournissent des LOD et des LOQ différentes. Une LOQ élevée ne signifie pas automatiquement que les données doivent être exclues.
94. Orientation pour l'inclusion/exclusion de données dans différents scénarios de LOQ/LOD
- Dans le cas où aucune LOQ/LOD n'est fournie pour un ensemble de données spécifique
 - Le pays/organisation ou l'observateur qui soumet les données doit être contacté dans un premier temps pour obtenir ces informations (c.-à-d. LOD et/ou LOQ).
 - Dans le cas où l'ensemble de données contient (presque) tous les résultats quantifiés: l'ensemble de données peut être utilisé⁴.
 - Lorsque l'ensemble de données contient une quantité significative de données censurées à gauche (c'est-à-dire des données individuelles sans valeurs quantifiées ou finies, généralement appelées données inférieures aux LOQ/LOD rapportées): l'ensemble de données ne doit pas être utilisé.
 - Dans le cas où une LOQ est fournie:
 - Identifier un seuil pour la LOQ dans l'analyse en fonction des LM considérées (exemples: $LOQ < LM$ en cours de discussion, $LOQ < 1/2 LM$ en cours de discussion).
 - Dans le cas où l'ensemble de données contient une quantité importante de données censurées à gauche: des informations contextuelles supplémentaires doivent être prises en compte (par exemple, les données proviennent-elles d'un pays importateur ou d'un pays producteur).
95. Si la quasi-totalité des données de l'ensemble de données sont inférieures à la LOQ ou déclarées comme non détectées (ND, <LOD), il n'est pas possible d'estimer des valeurs de centile élevées pour établir les LM proposées. Lorsqu'il n'y a qu'un petit nombre de valeurs quantitatives, l'ensemble de données doit être traité au cas par cas en suivant les conseils fournis dans la section «Traitement des ensembles de données comportant une grande proportion de données censurées à gauche». Dans ce cas, lorsque l'on propose des LM, il n'est pas approprié de calculer des valeurs de centile élevées en utilisant uniquement des valeurs quantitatives, car cela peut aboutir à des LM inutilement élevées.
96. Le document de travail du GTE devrait indiquer clairement les critères selon lesquels certaines données ont été exclues de l'ensemble de données en raison de LOD/LOQ élevées (par exemple, la LOD/LOQ rapportée pour certains échantillons est supérieure à la LM proposée, ou la LOD/LOQ rapportée pour certains échantillons est supérieure de «x» ordres de grandeur à la LOD/LOQ la plus basse de la grande majorité des autres échantillons de l'ensemble de données, ce qui suggère une erreur) ou si l'ensemble de données doit être exclu de l'analyse, car l'élimination de données individuelles peut introduire un biais.

Somme des composants et LOQ

97. Dans le cas de niveaux de contaminants qui sont une somme de composants (par exemple, les aflatoxines totales), les éléments suivants doivent être pris en compte.
- La règle générale est que les niveaux de contaminants qui sont une somme de composants sont rapportés comme la limite inférieure, c'est-à-dire que pour les niveaux non quantifiés de composants, les valeurs sont fixées à 0.
 - Les LOD ou les LOQ des composants individuels doivent être fournis pour les composants non quantifiés en dessous de la LOD ou de la LOQ.
 - Lorsque seules des données sur des composants individuels sont communiquées, les données individuelles peuvent être additionnées pour obtenir un résultat total.
 - Dans certains cas, il peut être approprié de déclarer les taux de contaminants qui sont une somme de composants en utilisant une approche de limite moyenne (c'est-à-dire que les niveaux non quantifiés des composants sont fixés à une $\frac{1}{2}$ LOQ (ou à la LOD si la LOQ n'est pas fournie) ou une approche de limite

⁴ Dans le modèle actuel de base de données GEMS/Aliments, le rapport de la limite de détection (LOD) ou de la limite de quantification (LOQ) n'est pas obligatoire en cas de résultats quantifiés. Par conséquent, ces données, rapportées à l'aide de ce modèle, pourraient être utilisées, notamment en cas d'insuffisance d'autres données, avec l'indication claire dans le document présenté par le GTE à la session plénière de la CCCF qu'aucune LOQ/LOD n'a été fournie pour ces données, conformément aux règles en vigueur au moment où les données ont été soumises.

supérieure (c'est-à-dire que les résultats non quantifiés sont fixés à 1 LOQ (ou à la LOD si la LOQ n'est pas fournie); toutefois, ces cas doivent être clairement identifiés à l'avance avant la soumission des données dans les instructions relatives à l'appel de données.

ANALYSE DES DONNÉES: GÉNÉRATION D'UN APERÇU DE DONNÉES

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

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

Couverture géographique des données d'occurrence fournies

99. Lors de l'évaluation d'une LM pour un produit de base particulier, l'ensemble de données doit inclure une représentation des régions de production qui sont importantes pour le commerce international. Par conséquent, il est utile (et peut être exigé à l'avenir) que le pays/la région qui fabrique les produits finis soit indiqué dans la base de données GEMS/Aliments (voir la section «Collecte et transmission des données»). De même, les principales régions de production pour certains produits peuvent être limitées géographiquement; si c'est le cas, le président du GTE peut rapporter cette information dans le document de travail. Dans ce contexte, les données des régions productrices doivent être prises en compte, car les données des pays qui importent les denrées alimentaires pourraient être biaisées si les denrées alimentaires doivent se conformer aux exigences du pays importateur, telles qu'une LM déjà établie dans ce pays. Si une région a des LM très restrictives, de sorte que les taux de contamination des aliments importés sont biaisés vers la gauche, il se peut que cela ne soit pas une représentation exacte de la variabilité de la contamination des régions productrices. Toutefois, les données des pays importateurs reflètent également les aliments (ingrédients) tels qu'ils sont échangés au niveau international et tels qu'ils sont consommés, et devraient être prises en compte dans une certaine mesure. En effet, une contamination supplémentaire pourrait avoir lieu pendant le transport depuis le pays producteur (par exemple, production de mycotoxines).
100. Dans certains cas, il pourrait être approprié de donner la priorité aux ensembles de données des pays producteurs par rapport aux ensembles de données des pays importateurs. Toutefois, dans ce cas, les ensembles de données des pays producteurs devraient refléter la mise en œuvre des bonnes pratiques agricoles et de fabrication (BPA, BPF) prévues dans les Code d'usages du Codex, si elles sont disponibles, et être représentatives des produits qui feraient l'objet d'un commerce international.
101. Une autre option consiste à fournir des analyses distinctes pour la production et l'importation d'ensembles de données par pays. Si possible, une analyse de sensibilité sur l'utilisation des données des pays producteurs par rapport aux pays importateurs devrait être réalisée pour guider la sélection des données pour l'élaboration de LM.
102. Ce n'est que s'il y a suffisamment de données indiquant des différences importantes dans les taux déclarés entre les régions ou entre les pays d'une région que l'analyse peut être effectuée par région ou par pays. Il convient de noter, dans le cas d'une approche par pays, que cette démarche doit être effectuée pour les principaux pays producteurs de la région et que des données suffisantes doivent être disponibles. (Voir la section sur l'Analyse statistique des données d'occurrence/Traitement des ensembles de données pour l'élaboration des LM).
103. En résumé, bien qu'il existe différentes approches possibles pour examiner les données des pays importateurs par rapport aux pays producteurs ou aux données régionales, il est important de garder à l'esprit que les LM du Codex sont des normes mondiales, et que l'approche par défaut pour le traitement des données devrait donc être d'analyser les données au niveau mondial.

104. Orientation pour les ensembles de données qui manquent de couverture géographique:
- Si la/les région(s) pour laquelle/lesquelles les données manquent est/sont une/des région(s) de production importante(s) et à condition qu'elle(s) s'engage(nt) clairement à fournir des données supplémentaires, quelques années supplémentaires (par exemple 2 à 3 ans) sont généralement accordées pour la collecte de données avant de poursuivre la discussion sur les propositions de LM, à condition qu'il n'y ait pas d'urgence et que le CCCF en convienne. Après l'expiration des années supplémentaires accordées, la discussion sur les LM pourrait être poursuivie sur la base des données disponibles, que la couverture géographique ait été atteinte ou non.
 - Si la ou les régions productrices importantes ne s'engagent pas à fournir les données supplémentaires ou si la collecte de données supplémentaires dans les délais convenus (par exemple, 2 à 3 ans) n'est pas possible, l'examen des LM se poursuivra sur la base des données disponibles ou sera interrompu.
 - Si la/les région(s) pour laquelle/lesquelles les données sont insuffisantes n'est/sont pas une/des région(s) de production importante(s): l'examen sur les LM se poursuivra sur la base des données disponibles.
 - S'il est urgent d'établir une LM pour la protection de la santé des consommateurs, un compromis doit être trouvé au sein de la CCCF pour fixer une LM sur la base des données disponibles. Dans ce cas, la LM peut être révisée dans un délai de 3 à 5 ans afin de déterminer si des ajustements sont nécessaires lorsque davantage de données seront disponibles.

Période couverte par les données d'occurrence fournies

105. Il convient que les données d'occurrence fournies se rapportent à plusieurs années de production pour l'élaboration de la LM. Les exigences peuvent être différentes selon les types de contaminants (par exemple, mycotoxines, toxines végétales, biotoxines marines, contaminants liés à la transformation et à l'environnement) et sont fonction de la variation annuelle supposée ou de l'évolution de la contamination dans le temps.
106. Pour les contaminants tels que les mycotoxines, dont on sait qu'ils varient d'une année à l'autre, les données des 10 dernières années peuvent fournir une très bonne représentation de la variation d'une année à l'autre; toutefois, dans certains cas, il peut être nécessaire de prendre en compte plus de 10 années de données (par exemple, si l'effort d'échantillonnage a été réduit au cours des dernières années ou si l'on dispose de moins d'ensembles de données de meilleure qualité). Pour d'autres contaminants, la variation d'une année sur l'autre est moins importante et des données plus récentes (ou une période plus courte) peuvent être sélectionnées. En tout état de cause, il convient de déterminer si les données datant de plus de 10 ans sont pertinentes pour l'analyse.
107. En outre, il pourrait être utile d'étudier/inclure des données plus anciennes pour savoir si certaines espèces/sous-groupes d'un groupe/catégorie d'aliments ont tendance à présenter des limites plus élevées.
108. Dans certains cas, il peut être approprié d'effectuer une analyse des tendances temporelles. Dans ces cas, les données de plus de 10 ans doivent être examinées pour déterminer si les concentrations ont changé/changent avec le temps et cela pourrait être utilisé pour déterminer si un certain nombre d'années de données devrait être utilisé pour l'élaboration des LM afin de représenter les concentrations actuelles.
109. Si un Code d'usages a été établi et mis en œuvre, les données examinées doivent porter sur les années suivant la mise en œuvre de ce Code d'usages afin de refléter les bonnes pratiques agricoles et de fabrication.
110. Si le GTE exclut des données sur la base de données collectées avant la mise en œuvre du Code d'usages et sans indication de la part d'un pays que les bonnes pratiques, tout au long de la chaîne de production, ont déjà été mises en œuvre avant l'établissement du Code d'usages, les exclusions et leur justification doivent être clairement documentées dans le document final présenté par le GTE à la session plénière du CCCF.

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

Considérations générales

111. Les sections suivantes expliquent les considérations à prendre en compte avant de procéder à une analyse statistique des données extraites/nettoyées, et la manière dont les résultats de l'analyse statistique doivent être présentés dans les GTE en vue de l'élaboration de LM applicables à l'échelle mondiale.
112. Les données extraites/nettoyées doivent faire l'objet d'une analyse statistique. Lors de l'évaluation des données, la première étape devrait consister à analyser la distribution de l'ensemble des données. En général, la distribution des données relatives aux contaminants dans les aliments tend à être asymétrique à droite, par exemple une distribution log-normale (voir Figure 1). Pour ces distributions non normales, l'utilisation de méthodes statistiques paramétriques, basées sur la distribution normale, n'est pas appropriée.

113. La CXS193 stipule à l'Annexe I que «les LM devraient être établies à un niveau (légèrement) supérieur au champ normal de variation des concentrations dans les produits de consommation humaine et animale». Cela signifie que pour élaborer une LM, il est nécessaire d'estimer des valeurs de centile élevées (généralement des valeurs du 95^e centile) avec un niveau de confiance élevé. En matière de sécurité alimentaire, un niveau de confiance de 95 % est généralement utilisé. La figure ci-dessous (Figure 1) explique, à l'aide d'une distribution modélisée, la relation entre une valeur de centile élevée, une LM hypothétique (généralement une valeur arrondie de la valeur de centile) et le pourcentage de points de données qui dépassent la LM proposée lorsque la LM est identique à la valeur du 98^e centile.

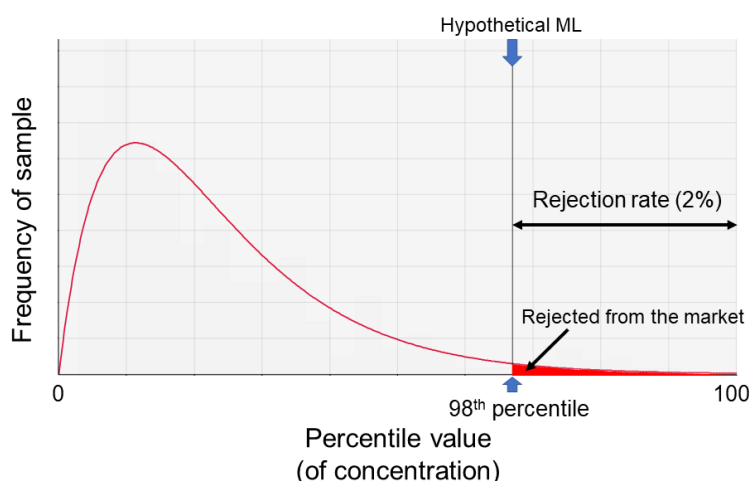


Figure 1. Représentation simplifiée de la relation entre une valeur de centile élevée, une LM hypothétique, le taux de rejet et le pourcentage d'échantillons dépassant la LM hypothétique.

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

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

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

Traitement des ensembles de données

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

115. Lorsque l'évaluation du Comité mixte FAO/OMS d'experts des additifs alimentaires (JECFA) suggère qu'un risque sanitaire lié à l'exposition à un contaminant est important, un nombre de points de données inférieur au nombre minimum de points de données (c'est-à-dire 59 points de données, voir le paragraphe 114 et l'Annexe II) serait considéré comme adéquat pour l'élaboration d'une LM, à condition que le niveau de confiance des valeurs de centile élevées estimées ne soit que légèrement inférieur au niveau de confiance élevé attendu, tel que 95 %. (Voir l'Annexe II pour un exemple de calcul du niveau de confiance) Par exemple, lorsqu'une LM est nécessaire de toute urgence pour protéger la santé des consommateurs, le GTE chargé de recommander des LM devrait envisager de recommander au CCCF l'élaboration de LM, même si seul un petit nombre de points de données est disponible. Si des données suffisantes sont disponibles à l'avenir, une révision des LM précédemment établies peut être envisagée.
116. S'il n'y a pas de risque sanitaire immédiat et que le nombre de points de données est insuffisant pour élaborer une LM, des appels de données supplémentaires peuvent être demandés. Toutefois, si, après plusieurs appels de données, le nombre de points de données disponibles reste très inférieur au nombre minimum requis, il convient de recommander au CCCF, au cas par cas, de décider d'élaborer une LM en utilisant l'ensemble de données limité ou d'interrompre les travaux. Une autre option peut être d'étendre la LM à un groupe alimentaire plus important, si une LM pour le groupe plus important est justifiée.
117. Pour les produits de base qui ne sont pas couramment consommés et/ou qui ne font pas l'objet d'un commerce international, la disponibilité des données d'occurrence peut être insuffisante. Dans de tels cas, le GTE chargé d'établir les LM devrait envisager de recommander au CCCF que la LM demandée pour la combinaison produit de

base/contaminant ne réponde pas aux critères décrits dans la norme CXS193 et le Manuel de procédure (Section 4.5 «Politique du Comité du Codex sur les contaminants dans les aliments en matière d'évaluation de l'exposition aux contaminants et aux toxines présents dans les groupes de denrées alimentaires ou d'aliments pour animaux») pour l'élaboration des LM.

118. Si le nombre de points de données est nettement inférieur au nombre minimal requis de points de données (c'est-à-dire 59, voir le paragraphe 114 et l'Annexe II) et qu'il n'y a pas de raison sérieuse d'élaborer immédiatement une LM, il n'est pas nécessaire d'effectuer d'autres analyses statistiques. Si des données supplémentaires sont nécessaires pour établir une LM statistiquement robuste, d'autres appels de données peuvent s'avérer nécessaires.
119. Lors de l'examen des LM existantes, même si seul un petit nombre de points de données provenant de régions limitées est disponible et/ou qu'aucune nouvelle donnée ne sera générée, le GTE chargé de recommander les LM ne devrait pas automatiquement recommander la révocation de la LM en raison du petit nombre de points de données, à moins que la LM ne soit incompatible avec les bonnes pratiques agricoles et de fabrication actuelles ou avec les données toxicologiques actuelles. S'il existe un risque potentiellement important lié à la consommation du produit de base, une option consisterait de recommander au CCCF de maintenir la LM existante, ou s'il n'y a plus de risque important pour la santé ou de barrière commerciale connue, une option consisterait de recommander au CCCF de révoquer la LM. Dans certains cas, lors de la révision d'une LM, s'il n'y a qu'un petit nombre de points de données pour un produit de base particulier, il peut être possible d'envisager de fusionner le produit avec le groupe d'aliments dont le produit a été exclu à l'origine (par exemple, en supprimant l'exclusion des conserves de *Brassica* d'une LM sur les conserves de légumes (réf. REP18/CF12, paragraphe 32)).

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

120. Même lorsque les points de données sont suffisants pour un groupe d'aliments entier, si les données sont séparées en fonction des aliments individuels dans ce groupe d'aliments, le nombre de points de données peut être trop faible pour l'élaboration de LM pour les aliments individuels. En général, l'analyse des données actuellement disponibles devrait commencer par l'élaboration d'un document de travail. Sur la base des données disponibles, le GTE devrait recommander une approche préliminaire de l'établissement des LM pour les aliments individuels et les groupes d'aliments, et recommander une formulation pour un nouvel appel de données. Si, après l'appel et la collecte de données, il s'avère qu'il y a moins de points de données disponibles qu'initialement prévu, il peut être nécessaire de modifier l'aliment ou les aliments ciblé(s) par la LM afin d'élargir l'éventail des aliments, par exemple en remplaçant un aliment par un sous-groupe d'aliments ou des sous-groupes d'aliments par des groupes d'aliments.
121. La pertinence de l'élaboration d'une LM pour un groupe d'aliments dépend de la similitude de la distribution des valeurs de concentration de contaminants dans les aliments individuels au sein du groupe d'aliments. Des tests statistiques non paramétriques, tels que le test U de Mann-Whitney (pour 2 ensembles de données) ou le test H de Kruskal-Wallis (pour 2 ensembles de données ou plus), peuvent être utilisés pour déterminer si la distribution des aliments du groupe peut être considérée comme similaire, même lorsque le nombre de points de données est relativement faible. Toutefois, comme les taux de contamination entre les produits d'un même groupe alimentaire peuvent varier de manière significative, il est important de faire preuve de souplesse dans le choix des méthodes et l'interprétation des tests statistiques. Si le nombre de points de données est relativement faible, il est également utile de comparer les ensembles de données à l'aide de diagrammes en boîte à moustaches, à condition que le pourcentage de données censurées à gauche soit inférieur à 25 % de l'ensemble de données concerné.
122. Si un aliment individuel présente une distribution des concentrations de contaminants différente de celle des autres aliments du groupe, il peut être nécessaire d'élaborer deux LM différentes, l'une pour le groupe d'aliments à l'exclusion de l'aliment individuel et l'autre pour l'aliment individuel. Des approches similaires peuvent être prises pour le(s) sous-groupe(s) du groupe d'aliments. Si les données relatives à une ou plusieurs denrées alimentaires sont insuffisantes pour atteindre le nombre minimal de points de données requis, des appels de données supplémentaires peuvent être lancés pour les aliments pour lesquels il est jugé nécessaire d'élaborer des LM. Si la consommation d'un aliment individuel dont la distribution diffère de celle du groupe d'aliments ne contribue pas de manière significative à l'exposition totale au contaminant en question, elle peut être considérée comme négligeable du point de vue de la protection de la santé du consommateur. Dans de tels cas, aucun appel de données supplémentaire n'est requis et son exclusion de l'élaboration d'une LM pour le groupe de denrées alimentaires peut être envisagée (par exemple, LM pour le plomb dans le sel, qualité alimentaire excluant le sel des marais (REP18/CF paragraphes 39-41). En ce qui concerne les groupes d'aliments et leurs sous-groupes, on peut se référer aux produits couverts par les normes de produits pertinentes du Codex, à la *Classification des denrées alimentaires et des aliments pour animaux* (CXA 4-1989) élaborée par le Comité du Codex sur les résidus de pesticides (CCPR) et au système de catégorisation des aliments utilisé par le Comité du Codex sur les additifs alimentaires (CCFA) pour les denrées alimentaires transformées.

123. Lors de l'élaboration d'une LM pour un groupe d'aliments plus large en raison de la disponibilité limitée de données pour des aliments ou sous-groupes individuels, certains aliments (ou sous-groupes) peuvent présenter des schémas de distribution différents de ceux d'autres aliments du même groupe d'aliments. Si les données ne sont pas suffisantes pour élaborer une LM distincte pour ces aliments, le GTE chargé de recommander des LM pourrait recommander au CCCF d'exclure ces aliments de l'élaboration de la LM jusqu'à ce que des données suffisantes soient disponibles.

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

124. Dans cette section, le terme «ensemble de données» fait référence à un ou plusieurs ensembles de données qui font partie des ensembles de données sélectionnés pour être utilisés dans le cadre de l'élaboration d'une LM. Cette section est particulièrement pertinente lorsque les ensembles de données utilisés pour l'élaboration des LM contiennent un taux élevé de points de données non quantifiés (par exemple, en raison de la faible sensibilité des méthodes d'analyse disponibles pour la concentration dans les échantillons, d'une fréquence d'occurrence extrêmement faible, etc.
125. Bien qu'aucune définition officielle du terme «censuré à gauche» ne figure dans les documents du Codex, en statistique, les données individuelles sans valeurs quantifiées (finies) sont appelées «données censurées à gauche», généralement désignées comme des données inférieures aux limites de quantification (LOQ) et aux limites de détection (LOD) rapportées.
126. Pour l'analyse statistique d'ensembles de données contenant des données censurées à gauche, il convient d'envisager des méthodes de substitution conventionnelles, en particulier pour le calcul de statistiques telles que le 95^e centile, ou pour l'estimation des taux de rejet d'échantillons et des réductions d'exposition pour les produits de base cibles dans le cadre de LM hypothétiques. Si l'ensemble de données contient un taux élevé de données censurées à gauche, l'analyse statistique utilisant uniquement des valeurs quantifiées n'est pas recommandée, car cette pratique introduit un biais dans les résultats de l'analyse statistique.

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

127. L'approche conventionnelle pour traiter les données censurées à gauche dans le cadre d'une analyse statistique consiste à utiliser un ou plusieurs des scénarios de substitution suivants:
- Scénario de l'estimation basse: les résultats inférieurs à la LQ sont remplacés par zéro, ou par la LOD si celle-ci est connue (les résultats <LOD sont remplacés par zéro);
 - Scénario de l'estimation haute: les résultats inférieurs à la limite de quantification sont remplacés par la valeur de la limite de quantification rapportée; et
 - Scénario de l'estimation moyenne (MB): Estimation ponctuelle entre les deux scénarios extrêmes (estimation basse et haute); on attribue alors une valeur égale à la LOQ/2, à la racine carrée de la LOQ, ou à (LOD + LOQ)/2 si la LOD est connue, pour les résultats d'analyse inférieurs à la LOQ rapportée. En général, la LOQ/2 est la plus utilisée.
- Pour chacun de ces scénarios, si la LOQ n'est pas rapportée et que seule la LOD l'est, utiliser la LOD comme alternative.
128. En général, en fonction de la distribution des données, ces méthodes de substitution peuvent être utilisées pour calculer des mesures de tendance centrale telles que la moyenne arithmétique lors du calcul de l'exposition alimentaire (voir Critères de santé environnementale 240 ou EHC 240⁵). Si le GTE chargé de recommander des LM doit effectuer des calculs préliminaires pour la réduction de l'exposition, le choix des scénarios d'estimation basse, moyenne ou haute (LB, MB ou UB) peut affecter la moyenne arithmétique calculée et l'exposition estimée pour les produits cibles sur la base de la moyenne arithmétique. Toutefois, pour l'élaboration de la LM, l'effet des valeurs censurées à gauche sur l'estimation empirique des valeurs de centile élevées peut être négligeable et il peut y avoir peu d'impact sur la LM quel que soit le scénario choisi, à moins qu'une grande majorité des points de données soient censurés à gauche (c.-à-d. <LOD).
129. Les ensembles de données comportant une grande proportion de données censurées à gauche doivent être traités au cas par cas, en fonction du résultat de l'évaluation des risques par le JECFA de la toxicité du contaminant et de la consommation de l'aliment concerné. Idéalement, les éléments d'estimation basse, haute ou moyenne devraient

⁵ CRITÈRES DE SANTÉ ENVIRONNEMENTALE 240, Principes et méthodes d'évaluation des risques liés à la présence de produits chimiques dans les aliments (OMS, 2009)

être calculés et présentés. Il est très important de connaître la distribution des valeurs quantifiées en cas de pourcentage élevé de données censurées à gauche lors de l'estimation des valeurs de centile élevées à l'aide d'une distribution modélisée pour élaborer des LM.

130. Lorsque la dispersion des valeurs quantifiées se situe dans une fourchette étroite (valeurs proches les unes des autres) et qu'elle est proche de la LOQ rapportée, l'élaboration d'une LM pourrait ne pas être nécessaire, à moins que le contaminant ne soit hautement toxique. Le GTE peut faire une recommandation au CCCF sur l'opportunité d'une LM dans cette situation.

Traitement de plusieurs ensembles de données

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

131. Les LM du Codex étant destinés à une application mondiale, ils devraient idéalement être basés sur des ensembles de données mondiaux. Bien que l'approche par défaut pour l'élaboration des LM consiste à utiliser l'ensemble des données mondiales combinées, les ensembles de données individuels par année ou par région sont fournis à titre d'information supplémentaire. Le CCCF doit décider au cas par cas si une LM doit être basée sur un ensemble de données mondiales ou sur un ensemble de données d'une région ou d'une année spécifique, après l'analyse statistique décrite dans cette section.
132. Des méthodes statistiques paramétriques sont disponibles pour comparer les schémas de distribution des ensembles de données individuels par région/pays ou par année. L'hypothèse nulle est que tous les ensembles de données sont supposés suivre la même distribution. Ces tests comprennent le test t (pour 2 ensembles de données) ou l'ANOVA (pour 2 ensembles de données ou plus).
133. De nombreux modèles de méthodes statistiques non paramétriques sont disponibles sur Internet. Parmi eux, des modèles MS Excel pour effectuer les tests U de Mann-Whitney (pour 2 ensembles de données), et les tests H de Kruskal-Wallis (pour 2 ensembles de données ou plus) peuvent être téléchargés sur le site web de la JMPR de la FAO⁶.
134. En outre, il est utile de tracer des diagrammes en boîte à moustaches ou des histogrammes pour chaque ensemble de données afin de comparer s'il existe des différences visuelles dans les distributions avant de combiner les ensembles de données. Il est préférable de ne dessiner un histogramme que lorsque l'ensemble de données contient un nombre suffisant de points de données (voir paragraphe 114 et Annexe II). Pour un ensemble de données avec un plus petit nombre de points de données, il est difficile de connaître la forme de la distribution à l'aide d'un histogramme, et un diagramme en boîte à moustaches est plus utile (voir Annexe III).
135. La proposition d'une ou plusieurs LM à l'aide d'un ou de plusieurs ensembles de données combinés (ensembles de données mondiales) a été faite de manière conventionnelle dans les GTE. Lorsqu'il n'y a pas de différence significative entre les distributions de plusieurs ensembles de données provenant de sources différentes, on considère qu'il est peu important d'effectuer des analyses et des comparaisons de données supplémentaires pour chaque ensemble de données individuel (bien qu'il serait idéal de le faire si les ressources et le temps le permettent).
136. Lorsque le nombre de points de données est très différent entre les ensembles de données individuels de différentes régions ou pays, l'ensemble de données combiné qui en résulte reflète principalement les conditions d'un pays ou d'une région ayant le plus grand nombre de points de données, plutôt que celles de tous les pays ou régions ayant fourni des données. Pour résoudre ce problème, il serait théoriquement possible, bien que cela nécessite un processus complexe, d'équilibrer les ensembles de données en les pondérant par le volume de production ou de commerce ou par tout autre facteur raisonnable. Toutefois, la méthodologie et la justification de l'utilisation de la pondération des données n'ont pas été examinées dans le passé par le CCCF et, en raison de sa complexité et de la charge de travail qui en découle, il n'est pas envisagé de le faire avant qu'une orientation pratique ne soit élaborée à l'avenir à cet effet. Si l'on craint une distorsion dans la distribution de l'ensemble de données combiné en raison de la présence d'un très grand ensemble de données, il convient d'examiner la question au cas par cas. Pour le moment, il peut être recommandé d'effectuer également des analyses pour chaque ensemble de données individuel, comme présenté dans les sections suivantes, ou de demander conseil au JECFA sur l'analyse des données.

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

137. Si un test statistique indique une différence significative entre la distribution de plusieurs ensembles de données,

⁶ «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/>

et que cette différence est importante, il est recommandé d'analyser les ensembles de données individuels en même temps que l'ensemble de données combiné pour l'élaboration de la LM. Toutefois, cette décision doit être prise au cas par cas, car l'ampleur des différences de distribution dépend généralement de la combinaison du produit de base et des contaminants examinés. Une justification de l'analyse des ensembles de données spécifiques séparément de l'ensemble de données combiné doit être fournie, et si aucune justification ne peut être trouvée, l'ensemble de données combiné doit être utilisé par défaut.

138. Lorsque l'on envisage d'utiliser des ensembles de données individuels, il est recommandé de comparer les résultats statistiques, tels que les valeurs de centile élevées des ensembles de données séparés, à ceux de l'ensemble de données combiné. Il convient de noter qu'il n'est pas possible d'obtenir des valeurs robustes de centile élevé pour les ensembles de données individuels ou combinés dont la taille des échantillons est inférieure au nombre minimum requis de points de données (voir paragraphe 114 et le Tableau 1 de l'Annexe II).
139. Il convient de noter que lorsque plusieurs ensembles de données sont examinés individuellement, plusieurs LM possibles peuvent être identifiées. Il n'entre pas dans le champ d'application de la présente Orientation de fournir des conseils sur le choix éventuel d'une LM, car c'est le CCCF qui décide de la LM appropriée, en tenant compte d'éléments qui peuvent varier d'un cas à l'autre.
140. Si les ensembles de données provenant de différentes régions/pays sont analysés séparément au moyen des méthodes statistiques recommandées dans le présent guide, le GTE chargé de recommander les LM pourrait fournir des éléments pour le choix de l'ensemble de données sur lequel fonder la LM pour examen et décision par le CCCF. Par exemple, si l'on a l'assurance que les ensembles de données présentant des concentrations élevées concernent des produits fabriqués selon de bonnes pratiques (Code d'usages du Codex ou BPA, BPF, etc.), il pourrait être utile alors de se concentrer sur les ensembles de données présentant des concentrations élevées afin d'envisager des LM applicables à l'échelle mondiale.

Réalisation d'une analyse des données par visualisation

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

Agrégation des données et calcul des statistiques descriptives

142. Les informations et les statistiques récapitulatives suivantes peuvent être présentées pour des ensembles de données d'occurrence répondant aux exigences minimales relatives:
 - Nombre de points de données totaux.
 - Nombre de points de données inférieurs aux LOQ (ou LOD) rapportées, et/ou rapport entre le nombre de données <LOQ (ou < LOD) sur le nombre total de points de données.
 - Fourchette des LOQ (LOD) parmi les données exclues ou incluses dans l'ensemble de données final utilisé pour recommander les LM.
 - Moyenne (moyenne arithmétique): si l'ensemble de données contient des points de données inférieurs à la LOQ (LOD), trois moyennes arithmétiques basées sur trois scénarios d'estimation basse, moyenne ou haute (LB, MB ou UB) peuvent être préparées (si la distribution est normale ou proche de la normale et symétrique).
 - Si la distribution est très asymétrique, une moyenne géométrique utilisant la même approche que ci-dessus pourrait être une option, mais elle n'a pas encore été utilisée au sein du CCCF.
 - Médiane (valeurs du 50e centile), mais si plus de 50 % des points de données sont inférieurs à la <LOQ, la médiane peut être indiquée comme «<LOQ» (ou <LOQ);
 - Valeurs de centile élevées (par exemple, 95^e, 97^e et 98^e centiles, si nécessaire, en fonction des discussions au sein du GTE sur le(s) taux de rejet approprié(s)); si plus de 95 %, 97 %, etc. des points de données sont inférieurs à la LOQ, les centiles associés pourraient être indiqués comme «<LOQ» (ou LOD);
 - Minimum.
 - Maximum: dans les cas où le maximum a été identifié comme une valeur aberrante potentielle et que la valeur maximale n'a pas été exclue de l'ensemble de données, il peut être utile de déclarer la 2e valeur la plus élevée, la 3e valeur la plus élevée, etc. pour un contexte supplémentaire.
 - Plage de données quantifiées.

- L'écart-type, qui est une mesure de la variation d'une distribution paramétrique.
 - Les valeurs interquartiles (voir Annexe III), qui sont une mesure de la variation d'une distribution non paramétrique.
143. Nombre de ces statistiques peuvent être facilement obtenues en utilisant les fonctions Excel, en utilisant un menu de statistiques descriptives dans les outils d'analyse de données de MS Excel, ou à partir de n'importe quelle autre application statistique. Les différentes applications statistiques utilisent des protocoles de calcul différents et renvoient donc des valeurs de centile différentes pour le même ensemble de données. Par conséquent, lors du calcul des valeurs de centile à l'aide d'applications informatiques, les valeurs obtenues doivent être soigneusement vérifiées par rapport aux fonctions utilisées et indiquer le nom et, le cas échéant, les détails de l'application utilisée pour le calcul.
144. Lorsque les données censurées à gauche constituent la majeure partie de l'ensemble de données, il peut être impossible de calculer des valeurs de centile élevées. Dans de tels cas, il est recommandé d'utiliser la méthode de substitution avec les scénarios d'estimation basse, haute ou moyenne. Bien que cela doive être fait au cas par cas, en fonction du nombre de valeurs quantifiées et de la distribution des contaminants, des méthodes telles que l'estimation des valeurs de centile élevées à partir des fonctions de densité de probabilité en modélisant la distribution des données d'occurrence peuvent être utilisées.

Calcul des taux de rejet à des LM hypothétiques

Estimation des LM hypothétiques

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

Calcul des taux de rejet à des LM hypothétiques

150. Le taux de rejet est défini par l'équation ci-dessous. Il peut être facilement obtenu en utilisant directement les fonctions de MS Excel (telles que la fonction COUNTIF) ou en utilisant des applications statistiques ou de modélisation/simulation après avoir modélisé chaque ensemble de données. Si une méthode différente est utilisée pour calculer le taux de rejet, elle doit être clairement indiquée dans le document de travail.
- $$\text{Taux de rejet (\%)} = (\text{nombre de points de données} > \text{LM hypothétique}) / (\text{nombre total de points de données}) \times 100$$
151. Il convient de noter que le taux de rejet peut être différent de celui prévu pour le centile élevé en raison de l'arrondissement. Plus le nombre de points de données utilisé pour calculer les taux de rejet est faible, plus l'incertitude dans l'estimation du taux de rejet est grande. Dans le calcul du taux de rejet, on suppose que les échantillons qui dépassent la LM hypothétique sont exclus du marché avec une probabilité de 100 % par l'application de la LM.

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

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

Amélioration du calcul des taux de rejet

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

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

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

157. Pour s'assurer que la LM proposée protège la santé des consommateurs, il pourrait être opportun d'évaluer quantitativement l'effet d'une LM hypothétique sur la réduction de l'exposition alimentaire du produit de base cible en comparant l'exposition sans LM et avec des LM hypothétiques. Dans le cas d'une révision d'une LM existante, l'exposition alimentaire du produit de base cible à la LM déjà établie est comparée à l'exposition aux nouvelles LM hypothétiques (LM révisées).

Une évaluation préliminaire de l'exposition au produit de base peut être réalisée en tant qu'information de référence dans les GTE chargés de recommander des LM si les ressources sont disponibles.

Si une évaluation détaillée ou plus complexe est nécessaire, (par exemple, l'évaluation de l'exposition globale), le CCCF peut demander au JECFA d'effectuer une évaluation des effets des LM hypothétiques sur la réduction du «risque lié à l'exposition alimentaire», car le calcul de l'exposition alimentaire est une fonction d'évaluation des risques dont devrait se charger le JECFA. Pour des exemples de calcul du taux de réduction de l'exposition alimentaire du produit cible, se référer à l'Annexe IV]

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

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

Préparation des recommandations finales au CCCF

159. Lors de la finalisation d'une analyse, le GTE chargé de recommander des LM peut être confronté à des questions difficiles telles que la nécessité d'une LM sur la base d'une évaluation préliminaire de l'exposition, la nécessité de plusieurs projets de LM proposés, l'ensemble de données à utiliser ou le taux de rejet approprié. Des perspectives différentes de celles envisagées lors de l'élaboration du mandat pour les travaux peuvent apparaître. Le GTE ne doit pas prendre de décision finale, mais préparer des questions et des recommandations à soumettre au CCCF dans son ensemble.
160. Si l'exposition alimentaire estimée dans le cadre de l'estimation supérieure est bien inférieure à la valeur indicative basée sur la santé (VABS), même en l'absence de LM, et qu'une LM proposée est égale ou proche de la valeur LOQ, la LM n'aurait qu'un faible impact sur la réduction de l'exposition alimentaire et ne serait pas nécessaire. Pour les contaminants pour lesquels aucune VABS n'a été établie, si toutes les données sont censurées à gauche, il pourrait être recommandé au CCCF d'établir la ou les LM à la valeur de la LOQ actuelle, s'il existe un risque potentiel pour la santé. Toutefois, si la plupart des données sont censurées à gauche et qu'il n'y a pas ou peu de problèmes de santé,

il convient de recommander au CCCF qu'il n'est pas nécessaire d'élaborer une LM. Par exemple, un ensemble de données combinées sur le plomb dans les œufs de poule frais après le nettoyage des données contenait 99 % de points de données censurés à gauche, allant de 0,001 à 0,257 mg/kg. La réduction de l'exposition calculée à des LM hypothétiques était faible, et la LM proposée se situait dans la fourchette des LOQ rapportées. Par conséquent, l'élaboration d'une LM pour le plomb dans les œufs de poule a été interrompue (réf. CX/CF 22/15/7).

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

PRESENTATION DES DONNEES DANS LES RAPPORTS DES GTE AU CCCF

Présentation de l'analyse des données: analyse statistique

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