

Application informatique : le logiciel CARAT

Patrice Buche, Lydie Soler, Jessica Tressou-Cosmao

► To cite this version:

Patrice Buche, Lydie Soler, Jessica Tressou-Cosmao. Application informatique : le logiciel CARAT. Analyse des risques alimentaires, Editions Tec

Doc Lavoisier, 400 p., 2006, 2743008946. <hal-01242338>

HAL Id: hal-01242338

<https://hal.archives-ouvertes.fr/hal-01242338>

Submitted on 11 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1.1. Application informatique : le logiciel CARAT

Patrice Buche, Lydie Soler et Jessica Tressou
INRA-Mét@risk

1.1.1. Introduction

De nombreux logiciels proposent des outils de calcul d'exposition et fournissent des estimateurs et des graphiques décrivant la distribution de l'exposition. Citons par exemple le projet MONTE CARLO financé par le programme Qualité de vie et ressources vivantes (Living Resources) du programme Santé, Alimentation et Nutrition du 5ème programme cadre de la commission européenne (<http://montecarlo.tchpc.tcd.ie/>, Gibney and van der Voet, 2003). Ce projet a donné lieu à la création de deux logiciels :

1. Le logiciel Monte Carlo Risk Assessment ou MCRA est développé par le RIKILT et Biometris (Université de Wageningen, Pays-Bas). Il permet à la fois l'évaluation des risques aigus et chroniques (Boer et al. 2005).
<http://mcra.rikilt.wur.nl/mcra/rsc/sjablonen/blauw/mcrahomenew.html>
2. Le Central Risk & Exposure Modelling e-solution ou CREME de l'IEFS (Institute of European Food Studies) et du Trinity Centre for High Performance Computing (Trinity College Dublin, Ireland) incorpore des procédures particulières pour traiter le risque lié aux migrants des emballages alimentaires. http://creme.tchpc.tcd.ie/contact_us/

Les problématiques propres à chacun des laboratoires de recherche impliqués ainsi que la difficulté à créer un outil unifié ont conduit chacune des équipes à développer son propre logiciel. Ces logiciels sont maintenant devenus des logiciels commercialisés et par voie de conséquence non disponibles sous la forme de licence open-source. Il n'est donc pas possible d'y intégrer de nouvelles fonctionnalités issues des méthodologies originales développées dans l'unité Mét@risk. En conséquence, Mét@risk a décidé de développer son propre logiciel baptisé CARAT (pour Chronic & Acute Risk Assessment). CARAT permet de rendre accessible certaines techniques de simulation à des non-statisticiens ainsi que les outils spécifiques développés au sein de l'unité. Les méthodes proposées dans les chapitres précédents (calcul déterministe ou non paramétrique de l'exposition, avec intervalles de confiance par bootstrap, Chapitres 3.1 et 3.2, modélisation de la censure des données de contamination, Chapitre 3.3) ainsi qu'un système d'aide au rapprochement des nomenclatures consommation et contamination sont donc disponibles via l'interface JAVA de CARAT.

La première section présente d'une part, les sources de données utilisées pour calculer l'exposition et d'autre part, les grandes étapes suivies par CARAT pour définir les paramètres d'une étude à partir de données provenant de ces sources. Dans la deuxième section, les deux modes de calcul de l'exposition proposés par CARAT sont présentés (déterministe et non paramétrique) ainsi que les statistiques produites par ces modes de calcul.

1.1.2. Préparation des données

L'objectif de CARAT est de fournir aux utilisateurs un outil d'évaluation de l'exposition des consommateurs à un risque alimentaire à partir d'un ensemble de données de consommations individuelles et d'un ensemble de données de contamination chimique. CARAT a été conçu afin d'être utilisable par des non statisticiens grâce à une interface fonctionnelle et conviviale.

Le calcul de l'exposition au risque est basé sur des modèles de calculs statistiques développés au sein de l'unité Mét@risk. Ces méthodes de calcul statistique ont été développées par Jessica Tressou dans le cadre de sa thèse intitulée *Méthodes Statistiques pour l'évaluation du risque alimentaire*.

Avant de calculer une exposition à un risque alimentaire, il faut définir le « jeu de données » sur lequel travailler. Ce travail préalable consiste en la création d'une étude. Une étude est composée de groupes d'individus appelés population cible, de groupes de produits et d'un contaminant chimique. Les différentes étapes du traitement sont présentées à travers une étude sur l'Ochratoxine A.

Présentation des données de consommation et de contamination

Les données en entrée d'un calcul sont de deux types, les données de consommation et les données de contamination. C'est à partir de ces données que l'on peut calculer l'exposition au risque. Les données sont présentées dans un format simplifié par soucis pédagogique, elles ne représentent pas exactement les structures utilisées dans l'implémentation.

Données de consommation

Dans sa version actuelle, CARAT utilise une base de données de consommations individuelles tirées d'une enquête AFSSA (Agence Française de Sécurité Sanitaire des Aliments) effectuée en 1999, appelée INCA (enquête Individuelle et Nationale sur les Consommations Alimentaires)¹. Cette étude de la consommation a été faite sur 3003 individus sur une durée d'une semaine. Dans la base de données de consommation, on dispose d'informations sur le consommateur (sexe, âge, poids, taille, catégorie socioprofessionnelle ...) et sur sa consommation individuelle (nature des repas, produits consommés, quantités consommées ...). Toute autre source de données de consommation individuelle peut être utilisée par CARAT à partir du moment où celle-ci peut être mise dans le même format que celui de la base INCA.

On peut représenter les données de consommation comme une matrice contenant les consommations des individus en produit ou en groupe de produits (association de plusieurs produits de même nature). Une ligne représente un individu et une colonne représente un groupe de produits.

¹ Ces données sont présentées dans le chapitre 1.2 et leur utilisation dans le cadre de l'évaluation du risque alimentaire est également traitée dans la section 3.1.2.

Matrice C des consommations pour n individus et p groupes de produits.

	1				p
1					
n					

$C_{i,j}$ = consommation de l'individu i pour le groupe de produits j .

Figure 1. Représentation schématique des données de consommation

Le niveau de granularité temporelle de la cellule (encore appelé fenêtre temporelle des données de consommation) dépend de la source de données de consommation. Avec la source de données INCA, trois niveaux sont possibles : consommation par semaine, par jour ou par occasion de repas. Une occasion de repas est une consommation prise lors d'un repas : petit déjeuner, déjeuner, dîner, encas.

Pour effectuer le calcul d'exposition, les données de consommation doivent être divisées par le poids corporel de l'individu. Cette transformation permet d'obtenir une exposition qui peut être comparée avec une valeur toxicologique de référence (voir Chapitre 1.3). Il s'agit dans l'étude Ochratoxine A de la DHT (Dose Hebdomadaire Tolérable) ou de la DJT (Dose Journalière Tolérable).

Tableau 1. Exemple d'une matrice de consommation avec une fenêtre temporelle hebdomadaire (données factices en g/kg/semaine) :

Groupes de produits								
Individus	Code individu	Fruits et légumes secs	Café	Produits à base de céréales	Jus de Fruit	Riz	Céréales	Viande
	4751701	13.33	0.00	40.25	13.75	10.42	18.75	4.17
	4751301	0.00	45.94	23.25	12.38	26.25	38.63	21.75
	4751101	0.00	0.00	64.09	0.00	18.04	43.57	5.27
	4749901	14.44	0.00	53.11	4.58	0.00	63.89	44.61
	4748601	0.00	0.00	40.47	0.00	11.33	43.87	0.00
	4748401	1.28	0.00	21.86	16.48	4.23	29.84	4.63
	4747101	0.00	0.00	37.88	6.00	15.00	28.41	2.10
	4747001	0.00	0.00	53.36	21.64	4.92	89.02	18.10
	4746001	4.50	0.00	47.40	42.80	30.00	54.00	34.80
	4745101	13.13	0.00	29.74	0.00	12.38	45.38	8.18
	4745001	6.67	0.00	25.94	16.67	3.61	39.22	5.92
	4744901	0.00	0.00	13.30	33.46	8.65	36.63	0.00
	4743601	8.82	0.00	50.65	56.47	8.82	38.82	9.26
	4743501	0.00	0.00	20.07	0.00	2.87	10.46	14.34
	4740401	3.75	29.53	5.55	0.00	8.63	45.84	7.99
	4740301	0.00	0.00	16.59	2.40	1.80	17.33	5.85
	4739401	0.00	0.00	26.71	0.00	2.88	27.12	10.10
	4738401	0.00	0.00	37.14	1.44	21.00	55.14	12.30
	4738201	11.76	0.00	2.79	0.00	23.82	3.75	6.62

Données de contamination

Les méthodes de calcul proposées dans CARAT peuvent être utilisées pour évaluer l'exposition d'une population d'individus à un constituant quelconque (contaminant, nutriment,...). Néanmoins, comme les applications qui ont été traitées jusqu'à présent concernent toutes des contaminants, nous utiliserons dans la suite de ce chapitre le terme de contaminant.

CARAT utilise la base de données de contamination développée au sein de l'équipe informatique de l'unité Mét@risk. Elle permet de stocker les mesures de contamination des aliments fournies par les partenaires de l'unité (DGAL,...) ou directement commanditées par Mét@risk. Cette base utilise le référentiel REGAL de l'équipe CIQUAL de l'AFSSA pour codifier les aliments. Les mesures de contamination sont intégrées dans la base de données par lots au fur et à mesure des apports. Un lot est un ensemble de mesures de contamination provenant d'un même partenaire à une date donnée.

Les données de contamination se présentent sous la forme de listes de mesures de contamination par groupe de produits. Le nombre de mesures est variable selon le groupe de produits. Une mesure de contamination est une donnée numérique si la donnée est précise (i.e. on a pu mesurer la contamination), sinon la mesure est dite censurée à gauche quand elle atteint les limites de quantification des techniques de mesures utilisées. Une mesure de contamination « censurée » à gauche peut être :

- inférieure à la limite de quantification (LOQ) qui peut être associée à la notion de « trace » (on sait qu'il y a une contamination mais on ne peut pas la quantifier)
- inférieure à la limite de détection (LOD) qui est le seuil en dessous duquel on ne peut pas détecter la présence du contaminant.

On a toujours $LOD < LOQ$. CARAT propose plusieurs traitements possibles des données censurées. Ils sont présentés dans la section suivante.

Dans l'exemple ci-dessous (Tableau 2), on voit que chaque groupe de produits peut avoir un nombre différent de mesures de contamination pour un même contaminant.

Tableau 2. Exemple de mesures de contamination d'un contaminant pour trois groupes de produits (données factices en µg/kg)

Fruits et légumes secs	1.2	1.8	<0.5	<1.0	1.9	0.9	1.2
Café	<0.5	<0.5	<0.5	<1.0			
Produits à base de céréales	1.5	2.3	<0.1	<0.5			

Ces différentes mesures proviennent des différents lots disponibles pour ce contaminant. Les mesures précédées d'un « < » sont censurées. Le groupe Produits à base de céréales a ainsi 50% de ses données censurées et le groupe Café 100%. Il suffit de comparer la mesure à la LOD ou la LOQ pour savoir si elle est inférieure à la LOD ou à la LOQ et procéder au traitement adéquat. Le traitement de la censure sera abordé dans la section suivante.

Définition des données dans CARAT

Dans ce paragraphe, nous expliquons comment créer des groupes de produits et d'individus à l'aide de l'interface de CARAT.

Groupes de produits

La notion de groupe de produits permet d'établir la correspondance entre une liste de noms de produits consommés, issus de la base de consommation, et une liste de noms de produits contaminés, issus de la base de contamination.

Cette étape de mise en correspondance est nécessaire car les deux bases de données n'ont pas le même référentiel (c'est-à-dire, pas les mêmes codes associés aux aliments).

Un groupe de produits est donc défini comme un ensemble de deux listes de produits de même nature. Par exemple, le groupe de produits Fruits et Légumes secs est composé d'une part d'une liste de produits consommés comportant entre autres les abricots secs, les pruneaux secs et les pistaches rôties salées et d'autre part d'une liste de produits contaminés comportant entre autres les abricots secs dénoyautés, les figues sèches et les amandes.

Description du groupe produit

Description Française		Description Anglaise	
Libellé	Fruits et légumes secs	Libellé	Dry fruit and vegetables
Description		Description	

Produits associés

Liste des produits consommés	Liste des produits contaminés
13001 - Abricot sec 13011 - Datte sèche 13013 - Figue sèche 13042 - Pruneau sec 13046 - Raisin sec 13051 - Apéritifs (fruits séchés pour apéritif) 15000 - Amande 15001 - Cacahuète 15002 - Cacahuète grillée salée 15003 - Châtaigne 15004 - Noisette 15005 - Noix 15007 - Noix de coco amande sèche 15008 - Noix du Brésil 15009 - Pistache rôtie salée 15010 - Sésame graine 15011 - Tournesol graine 15015 - Purée de marron en conserve 15016 - Crème de marrons vanillée en conserve 15019 - Noix de cajou salée	13001 - Abricot, sec, dénoyauté 13011 - Datte sèche, pulpe et peau 13012 - Figue, fraîche 13013 - Figue, sèche 13060 - Datte fraîche, pulpe et peau 13062 - Figue de Barbarie, pulpe sans graine 13063 - Figue de Barbarie, pulpe et graines 13081 - Datte Deglet-nour, pulpe et peau 13522 - Datte du désert, sèche, pulpe et peau 13523 - Datte du désert, fraîche, pulpe et peau 13524 - Datte naine, pulpe et peau 15000 - Amande 15001 - Cacahuète, Arachide 15002 - Cacahuète, grillée, salée 15004 - Noisette 15005 - Noix 15007 - Noix de coco, amande, sèche 15010 - Sésame, graine 15011 - Tournesol, graine 15033 - Noisette grillée

Mise à jour Annuler

Figure 2. Listes des produits consommés et des produits contaminés composant le groupe Fruits et légumes secs pour l'étude Ochratoxine A

Le choix des groupes de produits a un impact sur les statistiques qui seront calculées par CARAT. En effet, les calculs d'exposition sont effectués au niveau des groupes de produits. Lorsque l'on crée un groupe de produits, il faut donc aussi vérifier que l'on dispose d'un nombre suffisant de données de contamination. Plus il y a de données de contaminations disponibles, meilleure sera l'évaluation de l'exposition.

Dans CARAT, un groupe de produits est défini par un utilisateur et appartient à cet utilisateur qui est le seul à pouvoir le supprimer ou le modifier. La suppression est uniquement autorisée dans le cas où le groupe de produits n'est associé à aucune étude. Dans le cas contraire, il faut supprimer toutes les études associées à ce groupe de produits pour pouvoir le détruire.

Un utilisateur peut utiliser les groupes de produits définis par un autre utilisateur, ceci afin d'éviter de recréer ce qui existe déjà. Un groupe de produits peut ainsi appartenir à plusieurs études. On voit ainsi sur la Figure 3 ci-dessous que certains groupes définis pour l'étude Ochratoxine A ont aussi été utilisés dans d'autres études.

Liste des groupes de produits

☐ Afficher uniquement mes groupes de produits

Libellé	Description	Etudes associées	Createur	Date de creation	Liste des produits	Supprimer
Alcool	Boissons alcoolisée...	Test Etude Ochratoxine A	Utilisateur 1	24 juin 2005	Télécharger	
Café	Café	Test Etude Ochratoxine A	Utilisateur 1	24 juin 2005	Télécharger	
Céréales	Pains,Biscottes, au...	Etude Ochratoxine A	Utilisateur 1	24 juin 2005	Télécharger	
Fruits et légu...	Fruits et légumes s...	Etude Ochratoxine A	Utilisateur 1	24 juin 2005	Télécharger	
Jus de Fruit	Jus de fruit et fru...	Etude Ochratoxine A	Utilisateur 1	24 juin 2005	Télécharger	
Poissons	Poissons	Etude poisson methylmercure Test Mercure	Utilisateur 1	19 sept. 2005	Télécharger	
Produit à base...	Biscuits, Patisseri...	Etude Ochratoxine A	Utilisateur 1	24 juin 2005	Télécharger	
Riz	Riz semoule (+ à ba...	Etude Ochratoxine A	Utilisateur 1	24 juin 2005	Télécharger	
Viande	Abats viande(porc ...	Etude Ochratoxine A	Utilisateur 1	24 juin 2005	Télécharger	
Vin	Vins et boisson à b...	Etude Ochratoxine A	Utilisateur 1	24 juin 2005	Télécharger	

Figure 3. Liste des groupes de produits pour l'étude Ochratoxine A et leur utilisation dans d'autres études

Groupes d'individus

Un groupe d'individus représente un sous ensemble de la population de la base de consommation. Il est défini à partir de caractéristiques choisies par l'utilisateur. Par exemple, le groupe d'individus « Adultes » est défini comme l'ensemble des individus de plus de 15 ans.

La création des groupes d'individus repose sur la caractérisation de la population qui doit être réalisé en dehors de CARAT avec l'aide d'un outil d'interrogation de la base de données de consommation. Pour faciliter le travail, les utilisateurs disposent de requêtes Access matérialisant des vues prédéfinies sur la base de consommation. Nous avons par exemple, créé deux vues permettant de décrire d'une part chaque individu et d'autre part chaque ménage composant la population de la base de consommation.

Voici quelques unes des informations que chaque vue met à la disposition des utilisateurs :

- vue décrivant chaque individu : identifiant de l'individu, sexe, âge, poids, taille, indice de masse corporelle, profession, catégorie socioprofessionnelle, lien avec le chef de famille (chef de famille, enfant, conjoint...)
- vue décrivant chaque ménage : identifiant du ménage, identifiant de l'individu faisant partie du ménage, région d'habitation, nature de l'habitat (rural, agglomération parisienne...), revenu, situation du chef de famille.

Les utilisateurs peuvent ainsi sélectionner des individus sur la base d'une combinaison de ces critères et produire un fichier Excel contenant la liste des identifiants des individus sélectionnés. Ce fichier Excel servira de base pour la création du groupe d'individus dans CARAT.

Creation d'un groupe d'individu

Description Française		Description Anglaise	
Libellé	Adultes	Libellé	Adults
	Adultes		Adults
Description		Description	
Fichier des identifiants : C:\Lydie\MetaRisk\Livre analyse de risi		Parcourir...	
Créer Annuler			

Figure 4. Création du groupe d'individus Adultes

CARAT vérifie que la liste fournie contient bien des identifiants d'individus de la base de données de consommation et crée le groupe d'individus. Un individu peut faire partie de plusieurs groupes d'individus et un groupe d'individus possède au moins un individu. Les groupes d'individus se gèrent de la même manière que les groupes de produits en matière de création, suppression et d'utilisation. Un utilisateur peut ainsi utiliser les groupes d'individus définis par un autre utilisateur. Un groupe d'individus peut aussi être associé à plusieurs études.

Liste des groupes d'individus

☐ Afficher uniquement mes groupes d'individus

Libellé	Description	Nombre d'individus	Etudes associées	Createur	Date de creation	Fichier source	Supprimer
Adultes	Individus de plus d...	1985	Test Etude Ochratoxine A	Utilisateur 1	14 sept. 2005	Télécharger	
Enfants	Individus de moins ...	1018	Test Etude Ochratoxine A	Utilisateur 1	14 sept. 2005	Télécharger	

Figure 5. Deux groupes d'individus sont définis pour l'étude Ochratoxine A. Les adultes (individus de plus de 15 ans) et les enfants (individus de moins de 15 ans). Ils sont également utilisés dans l'étude Test.

Création d'une étude

Une fois les groupes de produits et individus créés, l'étude peut être à son tour créée (voir la Figure 6). Celle-ci concerne un seul contaminant, un ou plusieurs groupes d'individus et un ou plusieurs groupes de produits.

Création d'une étude

Description Française		Description Anglaise	
Libellé	Etude Ochratoxine A	Libellé	Ochratoxine A Study
	Etude Ochratoxine A		Ochratoxine A Study
Description		Description	
Créer Annuler			

Figure 6. Ecran pour la création d'une étude

Choix du contaminant

L'utilisateur doit choisir le contaminant sur lequel porte l'étude. Pour cela, il dispose de la liste de tous les contaminants disponibles dans la base de données de contamination. Après avoir sélectionné le contaminant, tous les lots de données stockés dans la base de données de contamination contenant des mesures associées à ce contaminant sont affichés. L'utilisateur choisit alors le ou les lot(s) qu'il souhaite utiliser pour l'étude.

Toutes les mesures d'un même contaminant ne sont pas toujours exprimées dans la même unité. Des facteurs de conversion entre unités de mesure ont été stockés en base afin d'exprimer toutes les mesures associées à une étude dans la même unité pour un même calcul.

Choix des groupes d'individus

L'utilisateur doit ensuite définir la population de son étude en sélectionnant, à l'aide d'une liste de cases à cocher, des groupes d'individus parmi ceux qui ont été créés auparavant. Par exemple, pour l'étude Ochratoxine A, les deux groupes Adultes et Enfants ont été sélectionnés. L'utilisateur peut sélectionner des groupes d'individus qu'il n'a pas lui-même créés. Afin de connaître la composition de chacun de ces groupes, il peut récupérer les fichiers Excel contenant la liste des identifiants composants chacun d'entre eux.

Création d'une étude

Description

Constituant

Population

Groupes de produits

Liste des groupes d'individus associés à l'étude :

Modifier la liste

Libellé	Description	Nombre d'individus	Createur	Date de creation	Fichier source
Adultes	Individus de plus d ...	1985	Utilisateur 1	14 sept. 2005	Télécharger
Enfants	Individus de moins ...	1018	Utilisateur 1	14 sept. 2005	Télécharger

Figure 7. Récapitulatif des groupes d'individus ajoutés à l'étude Ochratoxine A

Choix des groupes de produits

Pour définir les groupes de produits pour une étude, l'utilisateur choisit des groupes parmi ceux créés auparavant à l'aide d'une liste de cases à cocher. L'utilisateur peut sélectionner des groupes de produits qu'il n'a pas lui-même créés. Pour chaque groupe de produits, il peut télécharger le fichier Excel contenant la liste des produits contaminés et consommés composant ce groupe de produits.

Facteurs de conversion des données

Une fois les groupes de produits ajoutés à l'étude, il est possible de définir, pour chacun d'entre eux, des facteurs de « process » et de conversion.

Application d'un facteur de « process » sur les données de consommation

Les consommations utilisées pour le calcul sont données par groupe de produits. Il se peut qu'un produit donné ne corresponde pas à 100% au groupe de produits auquel il est rattaché.

Si on prend l'exemple du groupe de produits céréales, on trouve entre autres des produits comme le pain, le sandwich au jambon ou le croissant aux amandes. On peut considérer que la composition de ces aliments est la suivante :

- Un pain contient 100% de céréales,
- Un sandwich au jambon contient 80% de céréales et 20% de viande,
- Un croissant aux amandes contient 70% de céréales, 5% d'amandes et 15% d'autres ingrédients (beurre, sucre...).

Dans un groupe de produits donné, l'utilisateur peut ainsi spécifier un facteur de « process » (valeur comprise entre 0 et 100%) pour chaque produit (recette, mode préparation, etc.). Ce facteur indique la proportion du produit consommé qui peut être affectée au groupe de produits dont il fait partie.

Ainsi pour le groupe de produits céréales, les facteurs de process des aliments présentés précédemment peuvent être de :

- 100% pour le pain
- 80% pour le sandwich au jambon
- 70% pour le croissant aux amandes.

Si un produit apparaît dans plusieurs groupes de produits pour une même étude, il faut vérifier que la somme de ses facteurs de « process » est inférieure ou égale à 100%. Ainsi, le sandwich au jambon peut appartenir à deux groupes de produits : le groupe céréales et le groupe viande, avec par exemple un facteur de 80% pour le groupe céréales et un facteur de 20% pour le groupe viande.

Dans l'étude Ochratoxine A, nous avons décidé d'affecter la totalité de chaque produit au groupe dont il fait partie. Pour chaque groupe de produits, le facteur de « process » de chaque produit est fixé à 100%.

Conversion des mesures de contamination

Les facteurs de conversion permettent d'apporter des corrections sur les données de contamination lorsque les mesures disponibles en base ne correspondent pas exactement au contaminant étudié. Par exemple, il n'y a pas en base de mesures de contamination pour le méthylmercure. On peut alors utiliser les mesures du mercure disponibles en leur affectant un facteur de conversion comme proposé dans la section 3.1.5. Ce facteur est un pourcentage associé à un couple (groupe de produits, produit) dans une étude donnée.

Par exemple, le facteur de conversion du mercure en méthylmercure dans le groupe de produits de la mer, vaut 84% pour le poisson, 36% pour les crustacés et 43% pour les mollusques. Ces valeurs sont définies par un expert.

Pour l'étude Ochratoxine A, nous avons laissé le facteur de conversion à 100% pour tous les groupes de produits.

Transfert des données

La préparation des données pour l'étude Ochratoxine A est achevée. Les calculs à effectuer dans cette étude peuvent maintenant être créés. Auparavant, les données de consommation et de contamination sont transférées dans la base CARAT afin d'assurer la stabilité du jeu de données sur lequel sont effectués les calculs d'exposition. En effet, la base de données de contamination évolue au fur et à mesure de l'ajout de nouvelles données. Il est donc nécessaire de copier les données associées à une étude afin que chaque nouveau calcul de cette étude travaille toujours à partir du même jeu de données.

Le transfert des données se déroule en quatre étapes présentées dans la Figure 8 ci-dessous:

- transfert des noms des groupes de produits,
- transfert des noms des groupes d'individus,
- transfert des données de consommation, c'est-à-dire de l'ensemble des consommations des individus de la population définie en chacun des groupes de produits de l'étude,
- transfert des données de contamination, c'est-à-dire de l'ensemble des mesures de contamination effectuées sur les groupes de produits de l'étude.

La première étape consiste à importer les noms des groupes de produits définis pour une étude ainsi que les noms des produits de consommation et de contamination composant ces groupes de produits. Ils sont directement importés des bases de données de consommation et de contamination dans la base CARAT.

Les noms des groupes d'individus ainsi que les identifiants de chaque individu composant ces groupes sont ensuite importés dans la base CARAT (étape 2 sur la figure).

Les deux dernières étapes du transfert consistent à importer les données de consommation (étape 3) et de contamination (étape 4). Les données de consommation sont traitées afin d'être directement exploitables par les calculs. Ainsi, elles sont divisées par le poids corporel de chaque individu, les facteurs de processing définis leur sont appliqués et elles sont agrégées en fonction des groupes de produits définis par l'utilisateur. On procède ensuite à l'importation des données de consommation dans la base CARAT.

En ce qui concerne les données de contamination, chaque mesure n'est importée qu'une seule fois dans la base CARAT. En effet, quels que soient les groupes de produits choisis, les traitements à appliquer aux mesures de contamination (traitement de la censure, agrégation des données...) dépendent du calcul à effectuer et sont donc réalisés lors de l'exécution du calcul. Lors de la création d'une étude, on importe ainsi dans la base CARAT les mesures de contamination n'ayant jamais été transférées. On mémorise aussi l'association de chaque mesure de contamination à cette étude ce qui permet de ne pas prendre en considération les données qui pourraient être intégrées à la base de contamination par la suite.

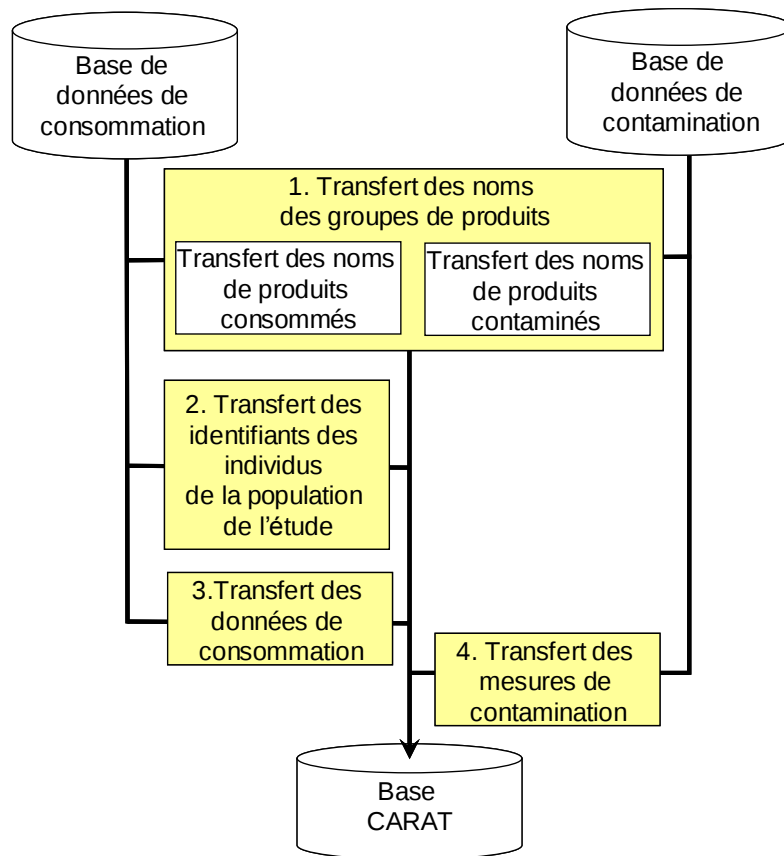


Figure 8. Etapes du transfert des données

Récapitulatif et choix des types de calculs statistiques

Nous avons décrit dans la section précédente, l'étape de préparation des données d'une étude. La Figure 9 illustre le récapitulatif qui clôt la création d'une étude.

Ce récapitulatif permet d'obtenir des informations synthétiques sur les données d'une étude :

- le nombre d'individus par groupe d'individus
- le nombre de mesures de contamination par groupe de produits
- le pourcentage de mesures de contamination censurées

Ces informations permettent de vérifier que l'on dispose d'un nombre de données suffisant pour effectuer les calculs statistiques. En particulier, le pourcentage de données censurées permettra de décider quel traitement de la censure doit être appliqué. Les différents traitements de la censure sont présentés dans les paragraphes suivants.

Description

Calculs déterministes

Calculs non paramétriques

Description Française

Libellé

Etude Ochratoxine A

Description

Etude Ochratoxine A

Description Anglaise

Libellé

Ochratoxine A Study

Description

Ochratoxine A Study

Cette étude porte sur le contaminant : **Ochratoxin A**

Liste des groupes de produit associés à l'étude

Libellé	Description	Consommation moyenne (g/semaine)	Nombre de mesures de contamination	Pourcentage de données censurées
Fruits et légu...	Fruits et légumes s...	115,886	24	100 %
Café	Café	925,451	12	100 %
Produit à base...	Biscuits, Pâtisseries...	696,018	10	100 %
Alcool	Boissons alcoolisées...	125,2	4	100 %
Vin	Vins et boisson à b...	394,422	910	77,033 %
Jus de Fruit	Jus de fruit et fru...	145,657	2	100 %
Riz	Riz semoule (+ à ba...	261,776	10	100 %
Céréales	Pains, Biscottes, au...	1 307,571	26	96,154 %
Viande	Abats viande(porc ...	237,691	659	8,346 %

Liste des groupes d'individus associés à l'étude

Libellé	Description	Nombre d'individus
Enfants	Individus de moins ...	1018
Adultes	Individus de plus d...	1985

Figure 9. Récapitulatif des paramètres de l'étude sur l'ochratoxine A

Dans CARAT, deux types de calcul ont été implémentés, le calcul déterministe et le calcul non paramétrique. Ces deux types de calcul prennent en entrée les données définies précédemment. Le résultat de base d'un calcul est l'exposition au risque chimique pour chaque individu de la population pour chaque groupe de produits défini. A partir de ces expositions, on peut calculer un certain nombre de statistiques. CARAT calcule un intervalle de confiance par bootstrap pour chaque statistique pour laquelle il est pertinent de le faire. Le calcul d'un intervalle de confiance par bootstrap est présenté dans la section suivante.

1.1.3. Calcul déterministe

Le calcul déterministe est effectué à partir des données de consommation hebdomadaire par groupe de produits. Il utilise les données de contamination disponibles pour chaque groupe de produits. Il s'effectue en trois étapes :

- le traitement des données censurées
- l'agrégation des données de contamination
- le calcul de l'exposition

Traitement des données censurées

CARAT effectue un certain nombre de traitements sur les données de contamination. Il traite la censure en remplaçant la mesure censurée par la LOD ou la LOQ multipliée par un coefficient.

Des traitements distincts sont prévus selon deux hypothèses :

Soient les paramètres α et β saisis par l'utilisateur et compris entre 0 et 1,

- Si la mesure est inférieure à la LOD, alors on remplace la mesure par αLOD
- Si la mesure est inférieure LOQ, alors on remplace la mesure par βLOQ ou $\alpha LOD + \beta LOQ$

Ceci permet de réaliser les substitutions recommandées par les comités d'experts internationaux, comme proposé dans le chapitre 3.2. Les 3 substitutions (H1, H2 et H3) proposées dans ce chapitre sont obtenues en choisissant respectivement $\alpha=1, 0.5$ ou 0 si la mesure est inférieure à la LOD et $\beta=1, 0.5$ ou 0 si la mesure est inférieure LOQ. De plus, une autre méthode de substitution courante consiste à remplacer les mesures inférieures à LOD par LOD/2 et les mesures inférieures à LOQ par (LOD+LOQ)/2 : le logiciel permet d'obtenir ce type de résultat en choisissant respectivement $\alpha = 0.5$ si la mesure est inférieure à la LOD et $\alpha = \beta = 0.5$ si la mesure est inférieure LOQ (cf. Counil et al, 2005).

On peut définir le même traitement de la censure pour tous les groupes de produits, ou alors définir un traitement différent pour chaque groupe de produits.

Par exemple, considérons les groupes de produits suivants :

Fruits et légumes secs	1.2	1.8	<0.5	<1.0	1.9	0.9	1.2
Café	<0.5	<0.5	<0.5	<1.0			
Produits à base de céréales	1.5	2.3	<0.1	<0.5			

On suppose que les LOD et les LOQ sont les mêmes pour toutes les mesures d'un même groupe de produits. On choisit de traiter ces groupes de produits avec les hypothèses suivantes :

- $\alpha = 1.0$ si la mesure est inférieure à LOD
- $\alpha = 0.5$ et $\beta = 0.5$ si la mesure est inférieure à LOQ
- Fruits et légumes secs : LOD = 0.5 et LOQ = 1.0
- Café : LOD = 0.5 et LOQ = 1.0
- Produit à base de céréales : LOD = 0.1 et LOQ = 0.5

On obtient alors les contaminations suivantes après traitement de la censure :

Fruits et légumes secs	1.2	1.8	0.5	0.75	1.9	0.9	1.2
Café	0.5	0.5	0.5	0.75			
Produits à base de céréales	1.5	2.3	0.1	0.30			

Agrégation des données de contamination

Après avoir traité la censure, les données de contamination sont agrégées par groupe de produits selon différentes méthodes :

- moyenne : on prend la moyenne de la liste des données de contamination,
- médiane : on prend la valeur médiane de la liste des données de contamination,
- percentile : on prend un percentile de la liste des données de contamination, le rang du percentile² est défini par l'utilisateur,
- maximum : on prend le maximum de la liste des données de contamination.

Le résultat du traitement des mesures de contamination est un vecteur de contamination par groupe de produits. En reprenant l'exemple précédent, on obtient les valeurs suivantes selon les différentes méthodes d'agrégation :

	Moyenne	Médiane	Percentile 95%	Maximum
Fruits et légumes secs	1.18	1.05	1.875	1.9
Café	0.56	0.5	0.71	0.75
Produits à base de céréales	1.05	0.90	2.18	2.3

Calcul de l'exposition

L'exposition au risque est obtenue en multipliant la matrice de consommation hebdomadaire par les mesures de contamination de chaque groupe de produits (agrégées selon une des méthodes décrites ci-dessus). Le résultat est l'exposition totale ventilée par groupe de produits d'un individu sur une semaine.

L'écran suivant (Figure 10) montre comment se passe la définition de l'ensemble des paramètres que nous venons de décrire.

La création d'un calcul déterministe passe par la définition des différents paramètres permettant le traitement de la censure, l'agrégation des données de contamination et le calcul de l'exposition.

Le traitement de la censure choisi pour le calcul déterministe de l'étude sur l'Ochratoxine A est identique pour tous les groupes de produits définis :

- si la mesure imprécise est inférieure à la LOD, elle est remplacée par $\alpha * LOD$ avec $\alpha=1.0$
- si la mesure imprécise est inférieure à la LOQ, elle est remplacée par $\beta * LOQ$, avec $\beta=1.0$

Le calcul déterministe pour l'étude Ochratoxine A utilise la moyenne comme méthode d'agrégation des mesures de contamination. Pour pouvoir calculer l'exposition hebdomadaire des individus, nous avons utilisé une DHT de 35 ng/kg/semaine.

Les intervalles de confiance et les percentiles sont présentés dans un prochain paragraphe.

² **Percentile** : On appelle k-ième percentile d'une distribution sur $\mathbb{R}$ la valeur x telle que les effectifs cumulés en x représentent k% des valeurs observées. On définit de même des quartiles, des déciles. On appelle interquartile la différence entre le troisième et le premier quartile.

Créer un nouveau calcul Déterministe

(A partir de l'étude Etude Ochratoxine A)

Description du calcul

	Description Française		Description Anglaise
Libellé	Calcul Moyenne	Libellé	Average calculation
Description	Calcul Déterministe Méthode d'agrégation: Moyenne	Description	Deterministic calculation Agrégation method: average

Définition de la DHT :

Saisir DHT à utiliser : ou choisissez la DHT à utiliser dans la liste :

La DHT doit être exprimée en ng/kg/semaine

Agrégation des mesures de contamination :

Méthode à utiliser : Rang du percentile (entre 0 et 100) :

Définition de la censure

☐ Définir une censure différente pour chaque groupe de produits

Groupe de produits	Hypothèse de censure si mesure < LOD	Hypothèse de censure si mesure < LOQ
Tous les groupes de produits	Remplacer par : $p * LOD$ $p = $ 1.0	Remplacer par : ☒ $p * LOQ ; p = $ 1.0 ☐ $p * LOD + p * LOQ ; p = $ $p = $

Calcul des percentiles sur l'exposition hebdomadaire des individus

Percentiles sur les consommateurs seuls

Rang du percentile (entre 0 et 100) : Ajouter

Percentile 69

Percentile 50

Percentile 98

Supprimer

Percentiles sur la population totale

Rang du percentile (entre 0 et 100) : Ajouter

Percentile 69

Percentile 50

Percentile 98

Supprimer

Calcul des intervalles de confiance :

- Nombre d'échantillons :

- Niveau de confiance (pourcentage) :

Figure 10. Création d'un calcul déterministe pour l'étude Ochratoxine A

1.1.4. Calcul non paramétrique

Le calcul non paramétrique est basé sur des tirages aléatoires avec remise dans les données de consommation et les données de contamination. Cette approche est plus réaliste que l'approche déterministe car dans la réalité les quantités consommées ne sont pas toujours contaminées de la même manière. Un calcul non paramétrique se décompose en trois étapes :

- le traitement des données de consommation
- le traitement des données de contamination
- le calcul de l'exposition

Traitement des données de consommation

Le calcul non paramétrique peut traiter des données de consommation selon différentes fenêtres temporelles (consommations par semaine, jour, ou occasion de repas) à partir du moment où les données disponibles le permettent. Il simule un nouvel échantillon de données en tirant aléatoirement avec remise des individus (et les consommations associées) dans la matrice des consommations. La matrice de consommation en entrée du calcul non paramétrique n'est donc pas la matrice de consommation initiale, mais une matrice de plus grande taille dans laquelle les consommations d'un individu peuvent apparaître plusieurs fois (ou pas du tout).

Le nombre de tirages M doit être supérieur au nombre d'individus dans la matrice de consommation. C'est une contrainte du modèle statistique qui nous assure des résultats consistants. Dans l'étude sur l'Ochratoxine A, la matrice de consommation comporte 3003 individus et 5000 tirages sont effectués.

Dans CARAT, l'utilisateur doit choisir :

- La fenêtre temporelle à utiliser (si la source de données le permet)
- Le nombre de tirages à effectuer dans les données.

Traitement des données de contamination

Le calcul non paramétrique traite aussi les données de contamination de manière aléatoire, mais seulement après avoir traité la censure des mesures de contamination. Dans ce cas, il existe deux types de traitements de la censure : le traitement « classique » qui est identique à celui du calcul déterministe, et le traitement utilisant la méthode de Kaplan Meier (outil de statistique non paramétrique).

Traitement « classique »

On procède de la même manière que le calcul déterministe mais on n'agrège pas les mesures de contamination. On obtient des vecteurs de mesures pour chaque groupe de produits dans lesquels CARAT tire aléatoirement avec remise M mesures.

Traitement de la censure avec Kaplan Meier

Dans le cas d'un traitement de la censure par la méthode de Kaplan Meier, on calcule pour chaque groupe de produits un estimateur de la fonction de répartition des mesures de contamination (fréquences cumulées). Celui-ci permettra de tirer aléatoirement une valeur parmi 0 et les valeurs non censurées observées.

Considérons le groupe de produits Fruits et légumes secs avec les valeurs suivantes :

Fruits et légumes secs	1.2	1.8	<0.5	<1.0	1.9	1.0	0.9	1.2
------------------------	-----	-----	------	------	-----	-----	-----	-----

On commence par calculer les valeurs distinctes non censurées, soit :

0.9	1.0	1.2	1.8	1.9
-----	-----	-----	-----	-----

Puis on calcule les probabilités d'être strictement inférieur à ces valeurs :

Probabilité d'être inférieur à 0.9	1/8= 0,125 (valeurs inférieures à 0.9 : <0.5)
Probabilité d'être inférieur à 1.0	3/8= 0,375 (valeurs inférieures à 1.0 : <0.5, 0.9, <1.0)
Probabilité d'être inférieur à 1.2	4/8= 0,5 (valeurs inférieures à 1.2 : <0.5, 0.9, <1.0, 1.0)
Probabilité d'être inférieur à 1.8	6/8=0.75 (valeurs inférieures à 1.8 : <0.5, 0.9, <1.0, 1.0, 1.2, 1.2)
Probabilité d'être inférieur à 1.9	7/8=0.875 (valeurs inférieures à 1.9 : <0.5, 0.9, <1.0, 1.0, 1.2, 1.2, 1.8)

Une probabilité est associée à chaque valeur distincte non censurée. Ensuite, pour choisir une mesure de contamination, une probabilité est tirée aléatoirement entre 0 et 1 pour trouver la mesure à utiliser. Si on tire une probabilité de 0.3 on prend la valeur de contamination 1. Si on tire une valeur inférieure à 0.125 on prend 0 comme valeur de contamination.

Il existe cependant un inconvénient dans cette méthode. On ne peut l'appliquer que s'il existe un nombre suffisant de mesures non censurées et une variabilité dans les valeurs de censure telle que certaines données quantifiées soient inférieures à une valeur de censure. Les mesures de contamination utilisées proviennent de différents laboratoires qui n'ont pas la même précision dans la quantification de la contamination. Certains laboratoires peuvent ainsi fournir des données quantifiées plus précises que les valeurs de censure d'autres laboratoires. Dans le cas où aucune donnée quantifiée ne serait inférieure à une valeur de censure, l'application de cette méthode non paramétrique est équivalente à la substitution par 0 des données censurées. Dans le cas contraire, on remplace la valeur de censure par la donnée quantifiée. En pratique, on retiendra un seuil de 30 mesures non censurées pour utiliser la méthode de Kaplan Meier. On traitera les groupes qui n'ont pas assez de mesures non censurées en tirant aléatoirement des mesures après un traitement de la censure par la méthode classique (coefficients sur la LOD ou la LOQ).

Calcul de l'exposition

Les deux entrées du calcul sont la matrice des consommations individuelles tirées aléatoirement et la matrice des données de contamination après traitement de la censure. Dans un calcul non paramétrique, l'utilisateur peut choisir la fenêtre temporelle (hebdomadaire, jour ou occasion de repas) ainsi que l'exposition (hebdomadaire ou journalière) du calcul.

Selon la fenêtre temporelle choisie, pour un individu et un groupe de produits donnés, on dispose :

- d'une donnée de consommation par semaine (fenêtre hebdomadaire),
- d'une donnée de consommation par jour (fenêtre jour),
- d'une donnée de consommation par occasion de repas (fenêtre occasion de repas).

Pour chacune de ces données de consommation, on tire aléatoirement une mesure de contamination (selon la méthode choisie). On multiplie la consommation par la mesure de contamination. Par conséquent, si les données de consommation sont exprimées au niveau jour, le calcul simule donc le fait que chaque consommation journalière est contaminée différemment. De la même manière, si les données sont exprimées par occasion de repas, le calcul simule le fait que chaque prise de l'aliment est contaminée de manière différente.

Le calcul de l'exposition dépend de la fenêtre temporelle choisie. Ainsi si une fenêtre temporelle hebdomadaire est choisie, seul un calcul d'exposition hebdomadaire est possible.

Si une fenêtre temporelle jour ou occasion de repas est choisie, un calcul d'exposition hebdomadaire ou journalière est possible.

Pour calculer une exposition, il faut agréger les données :

- au niveau semaine pour une exposition hebdomadaire
- au niveau jour pour une exposition journalière.

Le résultat d'un calcul non paramétrique est donc une exposition hebdomadaire ou journalière si la fenêtre temporelle le permet.

Les deux écrans suivant (Figure 11, Figure 12) nous montrent comment se passe le paramétrage d'un calcul non paramétrique.

Créer un nouveau calcul non paramétrique
(A partir de l'étude Etude Ochratoxine A)

Description du calcul :

Description Française	Description Anglaise
Libellé : Calcul censure Kaplan Meier Description : fenêtre temporelle: jour exposition hebdomadaire	Libellé : Kaplan Meier nonparametric calculation Description : time window: day weekly exposure

Fenetre Temporelle : Consommations par jour

Choix du type de calcul d'exposition : Calcul de l'exposition hebdomadaire

Définition de la DHT :

Saisir DHT a utiliser : 35 ou choisissez la DHT a utiliser dans la liste : ▼

La DHT doit être exprimée en ng/kg/semaine

Figure 11. Création d'un calcul non paramétrique (première partie)

Nombre de tirages : 5000

Définition de la censure : Censure Kanplan Meier

Vous devez saisir des hypothèses de censure pour tous les groupes de produits qui contiennent trop de mesures censurées

Groupe de produits	Pourcentage de données censurées	Hypothèse de censure si mesure < LOD	Hypothèse de censure si mesure < LOQ
Fruits et légumes secs	100 %	Remplacer par : $\beta * LOD$ $\beta = 1.0$	Remplacer par : ☒ $\beta * LOQ : \beta = 0.5$ ☐ $\beta * LOD + \beta * LOQ : \beta =$
Café	100 %	Remplacer par : $\beta * LOD$ $\beta = 1.0$	Remplacer par : ☒ $\beta * LOQ : \beta = 0.5$ ☐ $\beta * LOD + \beta * LOQ : \beta =$
Produit à base de céréales	100 %	Remplacer par : $\beta * LOD$ $\beta = 1.0$	Remplacer par : ☒ $\beta * LOQ : \beta = 0.5$ ☐ $\beta * LOD + \beta * LOQ : \beta =$
Alcool	100 %	Remplacer par : $\beta * LOD$ $\beta = 1.0$	Remplacer par : ☒ $\beta * LOQ : \beta = 0.5$ ☐ $\beta * LOD + \beta * LOQ : \beta =$
Jus de Fruit	100 %	Remplacer par : $\beta * LOD$ $\beta = 1.0$	Remplacer par : ☒ $\beta * LOQ : \beta = 0.5$ ☐ $\beta * LOD + \beta * LOQ : \beta =$
Riz	100 %	Remplacer par : $\beta * LOD$ $\beta = 1.0$	Remplacer par : ☒ $\beta * LOQ : \beta = 0.5$ ☐ $\beta * LOD + \beta * LOQ : \beta =$
Céréales	96,154 %	Remplacer par : $\beta * LOD$ $\beta = 1.0$	Remplacer par : ☒ $\beta * LOQ : \beta = 0.5$ ☐ $\beta * LOD + \beta * LOQ : \beta =$

Calcul des percentiles sur l'exposition hebdomadaire des individus :

Percentiles sur les consommateurs seuls

Rang du percentile (entre 0 et 100) : Ajouter

Percentile 69
Percentile 50
Percentile 98

Supprimer

Percentiles sur la population totale

Rang du percentile (entre 0 et 100) : Ajouter

Percentile 69
Percentile 50
Percentile 98

Supprimer

Calcul des intervalles de confiance :

- Nombre d'échantillons : 200

- Niveau de confiance (pourcentage) : 95.0

Valider Annuler

Figure 12. Création d'un calcul non paramétrique (suite)

Le premier paramètre à spécifier dans un calcul non paramétrique est la fenêtre temporelle. Comme nous l'avons expliqué précédemment, trois types de fenêtre temporelle sont possibles : consommations par occasion de repas, par jour et par semaine. En fonction de la fenêtre choisie, il est possible de choisir le niveau de granularité de l'exposition (par semaine ou par jour). La dose tolérable utilisée sera soit une dose hebdomadaire soit une dose journalière suivant les choix faits. Dans le calcul non paramétrique défini pour l'étude sur l'Ochratoxine A, une fenêtre temporelle journalière et un calcul d'exposition hebdomadaire ont été choisis. La dose tolérable utilisée pour le calcul de l'exposition a pour valeur 35 ng/kg/semaine.

Dans ce type de calcul, on effectue des tirages aléatoires avec remises pour générer les jeux de données sur lesquels vont porter les calculs. Il faut donc spécifier la taille de cet échantillon en indiquant le nombre de tirages à effectuer. Ici notre calcul effectuera 5000 tirages.

Ensuite, nous avons choisi d'effectuer un traitement de la censure par la méthode de Kaplan Meier pour les groupes de produits vins et viande qui disposent de suffisamment de données non censurées. Pour les autres groupes de produits, un traitement classique de la censure et identique pour chaque groupe de produits est effectué :

- si la mesure imprécise est inférieure à LOD, la mesure est remplacée par $\alpha * LOD$, avec $\alpha=1.0$
- si la mesure imprécise est inférieure à LOQ, la mesure est remplacée par $\beta * LOQ$, avec $\beta=0.5$

1.1.5. Construction d'intervalles de confiance (IC) par bootstrap

La technique de construction d'IC par bootstrap, présentée en détail dans le chapitre 3.1, permet d'associer un intervalle de confiance aux statistiques calculées. Cela permet de mesurer la variabilité des données. On réalise un certain nombre d'échantillons obtenus par tirage aléatoire d'observations de l'échantillon initial, appelé échantillon bootstrap. Sur chacun des échantillons bootstrap, on estime les différents paramètres du modèle statistique. On obtient par conséquent une distribution de paramètres.

Mode d'application du bootstrap

Cela consiste, dans un premier temps, à simuler un échantillon de données de consommation et de données de contamination grâce à un tirage aléatoire avec remise. Chaque échantillon contient le même nombre de données qu'au départ, mais certaines données peuvent être présentes plusieurs fois (ou pas du tout).

Par exemple, avec une matrice de consommation contenant 3000 individus, on tire aléatoirement 3000 individus avec remise pour obtenir une nouvelle matrice de taille identique dans laquelle on peut avoir plusieurs fois les mêmes consommations. On effectue le même traitement pour chaque liste de données de contamination pour obtenir de nouveaux échantillons de données.

Les données ainsi générées forment les nouvelles données en entrée d'un nouveau calcul (qu'il soit déterministe ou non paramétrique).

Cette suite d'opérations est effectuée N fois, et pour chaque itération on mémorise le résultat des statistiques calculées. On obtient alors une liste de valeurs pour chaque statistique qui nous permet de donner un intervalle de confiance sous la forme d'une borne inférieure et d'une borne supérieure.

Ces bornes sont en fait les percentiles aux extrémités de la distribution de chaque statistique. Par exemple, on prend le percentile 2.5% pour la borne inférieure et le percentile 97.5% pour la borne supérieure pour obtenir un intervalle de confiance à 95% (niveau de confiance de l'intervalle de confiance).

La Figure 13 schématise le plan d'exécution du calcul d'un intervalle de confiance par bootstrap.

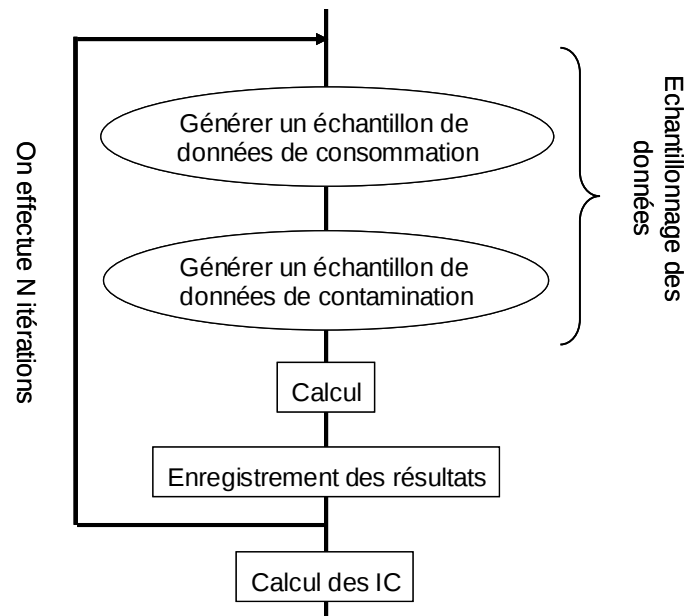


Figure 13 Description du plan d'exécution du bootstrap

Cependant, la technique du bootstrap n'est pas toujours valide, en particulier pour le calcul de l'intervalle de confiance de l'exposition maximum observée pour un groupe d'individus et pour l'IC de certains percentiles de l'exposition. La règle pratique pour les percentiles est la suivante : pour un percentile à p%, on calcule $\alpha = 1 - p/100$. Soit n le nombre d'individus traités pour un calcul, il faut impérativement que $\alpha * n$ soit « grand » au sens statistique. On retiendra un seuil de 30 en pratique.

Par exemple, pour un percentile à 99% ($\alpha = 0.01$), et pour une population de 3003 individus, on obtient $3003 * 0.01 = 30.03$. L'intervalle de confiance de ce percentile est donc valide, mais celui d'un percentile à 99.9% ne sera pas valide.

1.1.6. Résultats et sorties disponibles

Tous les types de calcul

Le résultat « de base » d'un calcul est l'exposition de chaque individu sur une semaine si la fenêtre temporelle choisie est hebdomadaire. Tous les calculs délivrent donc les résultats suivants:

- L'exposition sur une semaine pour chaque individu,
- L'amplitude de dépassement de la DHT pour chaque individu,
- L'exposition partielle de l'individu relative à un groupe de produits,
- La contribution de chaque groupe de produits dans l'exposition de l'individu.

Il faut ensuite donner des statistiques sur l'exposition des groupes d'individus, ces statistiques sont de deux sortes : les statistiques sur la population totale (tous les individus) et les statistiques sur les consommateurs seuls (on exclut les personnes qui ne consomment pas au moins un aliment contaminé). Pour ces deux catégories d'individus, on calcule :

- L'exposition moyenne,
- La contribution moyenne de chaque groupe de produits,
- L'écart type de l'exposition,
- L'exposition médiane,
- L'exposition maximum,
- Des percentiles sur l'exposition hebdomadaire (autant que l'utilisateur le désire),
- La probabilité de dépasser la DHT,
- La moyenne de l'amplitude de dépassement de la DHT.

CARAT calcule des intervalles de confiance par bootstrap (s'ils sont valides) pour chacune de ces statistiques.

Les résultats retournés aux utilisateurs sont donc toujours de la même forme et dépendent essentiellement de la fenêtre temporelle utilisée. On peut ainsi distinguer deux types de résultats : l'exposition hebdomadaire si la fenêtre temporelle choisie est la semaine, et l'exposition journalière si la fenêtre temporelle est par jour ou par occasion de repas.

Calcul sur des données détaillées au niveau jour ou occasion de repas

Un calcul peut être réalisé à partir de données de consommation détaillées (par jour ou par occasion de repas). Dans ce cas, on peut donner les statistiques supplémentaires suivantes :

- L'exposition par jour sur une semaine pour chaque individu,
- La probabilité de dépasser la DJT pour chaque jour,
- Le nombre de jours où l'individu dépasse la DJT.

Des statistiques sur ces résultats sont aussi demandées et devront être données pour la population totale et pour les consommateurs seuls :

- La moyenne de nombre de jours de dépassement de la DJT,
- Le pourcentage d'individus qui dépasse respectivement de 0 à 7 fois la DJT.

Un intervalle de confiance sera donné pour la moyenne du nombre de jours de dépassement.

Calcul déterministe

CARAT regroupe les résultats de chaque calcul (déterministe ou non paramétrique) dans un nouveau fichier Excel téléchargeable, permettant ainsi d'éventuels traitements ultérieurs. Ce fichier Excel est composé de sept onglets contenant les paramètres et les hypothèses du calcul ainsi que chacune des statistiques calculées.

Les résultats de l'exécution du calcul déterministe dont les paramètres ont été définis dans la Figure 10 sont présentés ci-dessous. Dans un calcul déterministe, seule la fenêtre temporelle hebdomadaire existe dans la version actuelle du logiciel, le calcul d'exposition effectué correspond donc à une exposition hebdomadaire.

Le premier onglet rappelle les hypothèses faites pour le calcul (voir Figure 14). On retrouve ainsi la méthode d'agrégation, la DHT et le niveau de confiance bootstrap choisis. Le deuxième onglet (Figure 15) rappelle les hypothèses de censure utilisées dans le calcul. Le troisième onglet (Figure 16) contient l'exposition hebdomadaire des individus ainsi que l'amplitude de dépassement de la DHT.

Le quatrième onglet (Figure 17) comporte la contribution de chaque groupe de produits dans l'exposition moyenne de l'ensemble de la population de l'étude. Le cinquième onglet (Figure 18) contient l'exposition hebdomadaire des individus pour chaque groupe de produits ainsi que la contribution de chaque groupe de produits dans cette exposition.

Les deux derniers onglets contiennent les statistiques sur les consommateurs seuls (Figure 19) et sur la population totale. Un consommateur seul est un individu de la population de l'étude ayant consommé au moins un produit des groupes de produits sélectionnés. On retrouve dans ces deux onglets les statistiques suivantes :

- la probabilité de dépasser la DHT,
- la moyenne de l'amplitude de dépassement de la DHT,
- l'exposition maximum,
- l'exposition moyenne,
- l'exposition médiane,
- l'écart type de l'exposition
- les percentiles sur l'exposition demandée.

Un intervalle de confiance est calculé pour chacune des statistiques de cet onglet, lorsque cela a un sens. Ainsi cet intervalle ne peut pas être calculé pour l'exposition maximum. En effet, si on tire des échantillons bootstrap, l'exposition maximum obtenue ne peut être qu'inférieure ou égale à l'exposition déjà obtenue. L'intervalle qui serait ainsi obtenu n'aurait aucun sens (par exemple, on dirait que 52 appartient à l'intervalle [28 ;42]).

Propriété	Valeur
Méthode d'agrégation	Moyenne
DHT utilisée	35.0
Niveau de confiance bootstrap	95.0

Figure 14. Rappel des hypothèses du calcul déterministe

Groupe de Produits	Censure si inférieur à LOD	Censure si inférieur à LOQ	
	Alpha	Aplha	Beta
Fruits et légumes secs	1,00	0,00	1,00
Café	1,00	0,00	1,00
Produit à base de céréales	1,00	0,00	1,00
Alcool	1,00	0,00	1,00
Vin	1,00	0,00	1,00
Jus de Fruit	1,00	0,00	1,00
Riz	1,00	0,00	1,00
Céréales	1,00	0,00	1,00
Viande	1,00	0,00	1,00

Figure 15. Rappel des hypothèses de censures (calcul déterministe)

Sur l'onglet présenté ci-dessous (Figure 16), on peut voir que l'individu 4751701 dépasse de 52.51 ng/kg/semaine la DHT qui vaut 35 ng/kg/semaine. Au contraire, l'individu 4749901 ne dépasse pas la DHT avec une exposition de 32.12 ng/kg.semaine.

	A	B	C	D	E	F	G	H
1	Exposition hebdomadaire des individus en ng/kg/semaine							
2								
3		Exposition	Amplitude de dépassement de la DHT					
4	4751701	87,51	52,51					
5	4751301	49,18	14,18					
6	4751101	27,46	0,00					
7	4749901	32,12	0,00					
8	4748601	81,11	46,11					
9	4748401	19,43	0,00					
10	4747101	30,23	0,00					
11	4747001	61,15	26,15					
12	4746001	53,39	18,39					
13	4745101	32,28	0,00					
14	4745001	47,56	12,56					
15	4744901	32,35	0,00					
16	4743601	39,72	4,72					
17	4743501	60,02	25,02					
18	4740401	86,27	51,27					
19	4740301	29,19	0,00					
20	4739401	40,04	5,04					
21	4738401	48,40	13,40					
22	4738201	26,04	0,00					
23	4738101	35,70	0,70					

Figure 16. Exposition hebdomadaire des individus (calcul déterministe)

Les groupes de produits Céréales et Produits à base de céréales sont les deux groupes de produits contribuant pour plus de 70% dans l'exposition moyenne (voir la Figure 17).

	A	B	C	D	E	F	G	H
1	Contribution de chaque groupe de produits dans l'exposition moyenne							
2								
3			Intervalle de confiance					
4	Groupe de produits	Contribution	Borne inférieure	Borne Supérieure				
5	Fruits et légumes secs	3,61%	3,43%	3,83%				
6	Café	3,18%	2,99%	3,34%				
7	Produit à base de céréales	24,00%	23,40%	24,51%				
8	Alcool	0,42%	0,36%	0,48%				
9	Vin	3,07%	2,85%	3,35%				
10	Jus de Fruit	4,05%	3,81%	4,36%				
11	Riz	8,52%	8,10%	8,90%				
12	Céréales	46,69%	46,08%	47,32%				
13	Viande	6,45%	6,24%	6,67%				
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								

Figure 17. Contribution de chaque groupe de produits dans l'exposition moyenne (calcul déterministe)

L'onglet ci-dessous montre l'exposition de chaque individu par groupe de produits et la contribution de ces groupes de produits dans l'exposition totale de chaque individu. L'individu 4751701 a une exposition de 54,79 ng/kg/semaine pour le groupe Produits à base de céréales ce qui représente 62,61% de son exposition totale.

Microsoft Excel - Calcul Moyenne-resultats.xls

Exposition hebdomadaire des individus pour chaque groupe de produits (avec contribution) en ng/kg/semaine

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1																				
2																				
3		Fruits et légumes secs		Café		Produit à base de céréales		Alcool		Vin		Jus de Fruit		Riz		Céréales		Viande		
4		Exposition	Contribution	Exp.	Contrib.	Exposition	Contribution	Exp.	Contrib.	Exp.	Contrib.	Exposition	Contribution	Exp.	Contrib.	Exp.	Contrib.	Exposition	Contribution	
5	4751701	2,50	2,86%	0,00	0,00%	54,79	62,61%	0,00	0,00%	0,00	0,00%	0,00	0,00%	5,25	6,00%	18,12	20,71%	6,85	7,82%	
6	4751301	0,00	0,00%	0,00	0,00%	22,46	45,66%	0,00	0,00%	0,00	0,00%	0,00	0,00%	2,08	4,24%	16,24	33,02%	8,40	17,09%	
7	4751101	0,00	0,00%	0,00	0,00%	14,38	52,37%	0,00	0,00%	0,00	0,00%	0,43	1,56%	2,57	9,37%	8,85	32,24%	1,23	4,47%	
8	4749901	2,17	6,77%	0,00	0,00%	18,16	56,54%	0,00	0,00%	0,00	0,00%	0,00	0,00%	2,17	6,77%	8,72	27,15%	0,89	2,77%	
9	4748601	0,00	0,00%	0,00	0,00%	74,71	92,11%	0,00	0,00%	0,00	0,00%	0,67	0,82%	0,00	0,00%	2,92	3,60%	2,81	3,46%	
10	4748401	0,00	0,00%	0,00	0,00%	6,13	31,54%	0,00	0,00%	0,00	0,00%	0,00	0,00%	1,79	9,19%	9,76	50,23%	1,76	9,04%	
11	4747101	2,17	7,19%	0,00	0,00%	9,47	31,32%	0,00	0,00%	0,00	0,00%	0,00	0,00%	0,00	0,00%	15,84	52,38%	2,76	9,11%	
12	4747001	0,67	1,09%	0,00	0,00%	15,98	26,13%	0,00	0,00%	0,00	0,00%	6,75	11,04%	6,25	10,22%	29,96	49,00%	1,54	2,52%	
13	4746001	0,00	0,00%	0,00	0,00%	22,00	41,20%	0,00	0,00%	0,00	0,00%	2,28	4,26%	1,52	2,84%	23,80	44,58%	3,80	7,11%	
14	4745101	0,00	0,00%	0,25	0,77%	20,16	62,46%	0,00	0,00%	0,00	0,00%	0,39	1,22%	1,32	4,08%	9,08	28,13%	1,08	3,33%	
15	4745001	1,25	2,63%	0,00	0,00%	15,48	32,55%	0,00	0,00%	0,00	0,00%	5,00	10,51%	0,00	0,00%	21,51	45,24%	4,31	9,07%	
16	4744901	0,24	0,75%	0,00	0,00%	12,07	37,32%	0,00	0,00%	0,00	0,00%	4,02	12,44%	1,22	3,77%	10,55	32,62%	4,24	13,10%	
17	4743601	0,00	0,00%	0,40	1,00%	16,87	42,47%	0,00	0,00%	0,00	0,00%	0,00	0,00%	1,52	3,81%	18,34	46,17%	2,60	6,55%	
18	4743501	0,00	0,00%	0,00	0,00%	5,50	9,16%	0,00	0,00%	0,00	0,00%	0,00	0,00%	7,80	13,00%	46,72	77,84%	0,00	0,00%	
19	4740401	1,20	1,39%	0,00	0,00%	21,09	24,44%	0,00	0,00%	0,00	0,00%	21,00	24,34%	3,20	3,71%	29,02	33,64%	10,76	12,47%	
20	4740301	0,00	0,00%	0,00	0,00%	12,10	41,45%	0,00	0,00%	0,29	1,00%	0,00	0,00%	6,11	20,94%	7,33	25,10%	3,36	11,52%	
21	4739401	0,00	0,00%	0,52	1,30%	15,42	38,51%	0,00	0,00%	0,12	0,31%	0,00	0,00%	6,19	15,46%	16,42	41,02%	1,36	3,40%	
22	4738401	0,00	0,00%	0,00	0,00%	14,71	30,39%	0,00	0,00%	0,00	0,00%	3,18	6,57%	0,00	0,00%	25,51	52,71%	5,00	10,33%	
23	4738201	1,35	5,19%	0,21	0,80%	12,46	47,84%	0,00	0,00%	0,00	0,00%	0,00	0,00%	0,00	0,00%	9,88	37,93%	2,14	8,23%	

Expositions hebdo par groupe / Stats consommateurs seuls / Stats population totale

Prêt

Somme=88,51462104

NUM

Figure 18. Exposition hebdomadaire des individus pour chaque groupe de produits et contribution des groupes de produits dans l'exposition des individus (calcul déterministe)

Statistiques sur les consommateurs seuls en ng/kg/semaine			
Nom	Valeur	Intervalle de confiance	
		Borne inférieure	Borne Supérieure
5 Probabilité de dépasser la DHT	36,33%	34,73%	38,20%
6 Moyenne de l'amplitude de dépassement de la DHT	8,31	7,76	8,94
7 Exposition Maximum	209,53	Bootstrap non valide	
8 Exposition Moyenne	35,09	34,33	35,98
9 Exposition Médiane	28,76	28,10	29,58
10 Ecart type l'exposition	22,73	21,78	23,83
Percentiles sur l'exposition			
Rang	Valeur	Intervalle de confiance	
		Borne inférieure	Borne supérieure
15 95.0	81,54	77,84	85,36
16 90.0	64,78	62,63	67,66
17 50.0	28,76	28,10	29,58

Figure 19. Statistiques sur les consommateurs seuls (calcul déterministe)

Dans notre étude sur l'Ochratoxine A, les statistiques obtenues pour les consommateurs seuls et la population totale sont identiques. En effet, tous les individus consomment au moins un produit des groupes de produits sélectionnés.

Calcul non paramétrique

Les résultats de l'exécution du calcul non paramétrique dont les paramètres ont été définis dans les Figure 11 et Figure 12 sont présentés ci-dessous. Le fichier Excel contenant les résultats du calcul non paramétrique comporte sept onglets contenant les paramètres et les hypothèses du calcul ainsi que chacune des statistiques calculées. Nous avons choisi pour ce calcul non paramétrique une fenêtre temporelle par jour et un calcul de l'exposition hebdomadaire.

Le premier onglet du fichier Excel (Figure 20) rappelle les paramètres du calcul à savoir le type de censure, la DHT utilisée (ou DJT suivant la fenêtre temporelle choisie), le nombre de tirages effectués dans les données de consommation et de contamination ainsi que le niveau de confiance des intervalles de confiance calculés par bootstrap. Le deuxième onglet (Figure 21) rappelle les hypothèses de censure. Si le traitement de la censure par la méthode de Kaplan Meier a été choisi, on ne retrouve dans cet onglet que les groupes de produits pour lesquels on ne peut pas appliquer cette méthode. Dans le cas d'un traitement classique de la censure, on retrouve les hypothèses de censure choisies pour tous les groupes de produits de la même manière que pour le calcul déterministe.

Le troisième onglet (Figure 22) contient l'exposition hebdomadaire des individus ainsi que l'amplitude de dépassement de la DHT (ou DJT suivant la fenêtre temporelle choisie). Le quatrième onglet (Figure 23) présente la contribution de chaque groupe de produits dans l'exposition moyenne. Un intervalle de confiance par bootstrap est calculé pour chacune de ces contributions. Dans le cinquième onglet (Figure 24), on trouve l'exposition hebdomadaire des individus pour chaque groupe de produits ainsi que la contribution des groupes de produits dans l'exposition des individus.

Les deux derniers onglets présentent les statistiques sur les consommateurs seuls (Figure 25) et sur la population totale. On retrouve dans ces deux onglets les statistiques suivantes :

- la probabilité de dépasser la DHT,
- la moyenne de l'amplitude de dépassement de la DHT,
- l'exposition maximum,
- l'exposition moyenne,
- l'exposition médiane,
- l'écart type de l'exposition
- les percentiles sur l'exposition demandée.

Un intervalle de confiance par bootstrap est calculé pour chacune de ces statistiques lorsque cela a un sens.

Propriété	Valeur
Type de censure	Censure Kaplan Meier
DHT utilisée	35.0
Nombre de tirages	5000
Niveau de confiance bootstrap	95.0

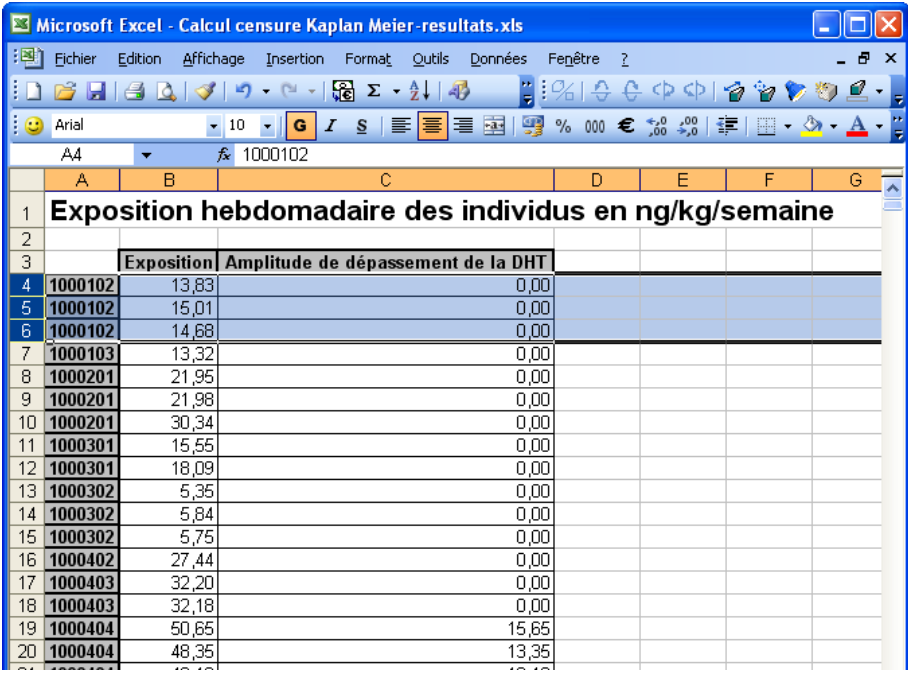
Figure 20. Rappel des paramètres du calcul non paramétrique

L'étude Ochratoxine A comprend neuf groupes de produits. Dans la Figure 21 ci-dessous, on retrouve les sept groupes de produits ne disposant pas d'assez de mesures de contamination non censurées. Par contre les groupes de produits Vin et Viande ne sont pas présents sur cette figure car la méthode de Kaplan Meier peut être appliquée.

Groupe de Produits	Censure si inférieur à LOD	Censure si inférieur à LOQ	
	Alpha	Aplha	Beta
Fruits et légumes secs	1,00	0,00	0,50
Café	1,00	0,00	0,50
Produit à base de céréales	1,00	0,00	0,50
Alcool	1,00	0,00	0,50
Jus de Fruit	1,00	0,00	0,50
Riz	1,00	0,00	0,50
Céréales	1,00	0,00	0,50

Figure 21. Rappel des hypothèses de censures (calcul non paramétrique)

Sur la Figure 22 ci-dessous, on peut voir que l'individu 1000102 apparaît plusieurs fois. Cela vient des tirages avec remise à partir des données de consommation.



Exposition hebdomadaire des individus en ng/kg/semaine						
	Exposition	Amplitude de dépassement de la DHT				
1000102	13,83	0,00				
1000102	15,01	0,00				
1000102	14,68	0,00				
1000103	13,32	0,00				
1000201	21,95	0,00				
1000201	21,98	0,00				
1000201	30,34	0,00				
1000301	15,55	0,00				
1000301	18,09	0,00				
1000302	5,35	0,00				
1000302	5,84	0,00				
1000302	5,75	0,00				
1000402	27,44	0,00				
1000403	32,20	0,00				
1000403	32,18	0,00				
1000404	50,65	15,65				
1000404	48,35	13,35				

Figure 22. Exposition hebdomadaire des individus (calcul non paramétrique)

Les groupes de produits Céréales et Produits à base de céréales contribuent pour près de 72% dans l'exposition moyenne. Lors du calcul déterministe, cette contribution vaut 70%. La différence vient du tirage aléatoire effectué dans le calcul non paramétrique : on peut tirer plusieurs fois le même individu, ce qui influe sur l'exposition.

Contribution de chaque groupe de produits dans l'exposition moyenne			
Groupe de produits	Contribution	Intervalle de confiance	
		Borne inférieure	Borne Supérieure
Fruits et légumes secs	4,09%	3,71%	4,41%
Café	3,67%	3,29%	3,88%
Produit à base de céréales	24,56%	23,34%	25,59%
Alcool	0,37%	0,32%	0,47%
Vin	2,67%	1,99%	3,00%
Jus de Fruit	4,44%	4,04%	4,94%
Riz	9,28%	8,92%	9,97%
Céréales	47,15%	45,49%	49,84%
Viande	3,78%	1,93%	5,95%

Figure 23. Contribution de chaque groupe de produits dans l'exposition moyenne (calcul non paramétrique)

Microsoft Excel - Calcul censure Kaplan Meier-resultats.xls

<

Figure 24. Exposition hebdomadaire des individus pour chaque groupe de produits et la contribution de chaque groupe de produits dans l'exposition individuelle (calcul non paramétrique)

Statistiques sur les consommateurs seuls en ng/kg/semaine			
Nom	Valeur	Intervalle de confiance	
		Borne inférieure	Borne Supérieure
Probabilité de dépasser la DHT	31,12%	28,36%	34,77%
Moyenne de l'amplitude de dépassement de la DHT	6,73	5,64	7,67
Exposition Maximum	210,37	Bootstrap non valide	
Exposition Moyenne	31,73	30,19	33,77
Exposition Médiane	25,71	24,43	27,66
Ecart type l'exposition	21,49	19,58	22,66
Percentiles sur l'exposition			
Rang	Valeur	Intervalle de confiance	
		Borne inférieure	Borne supérieure
95.0	73,63	69,48	79,56
90.0	60,13	55,50	63,68
50.0	25,71	24,43	27,66

Figure 25. Statistiques sur les consommateurs seuls (calcul non paramétrique)

Dans notre étude sur l'Ochratoxine A, les statistiques obtenues pour les consommateurs seuls et la population totale sont identiques. En effet, tous les individus consomment au moins un produit des groupes de produits sélectionnés.

1.1.7. Conclusion et perspectives

CARAT est une application Web développée en utilisant les technologies Java/J2EE dont l'architecture est décrite à la Figure 26. Le dialogue avec l'utilisateur s'effectue à partir d'un navigateur Web standard comme Mozilla Firefox, Internet Explorer, Les pages de dialogue sont générées en HTML par l'application CARAT par l'intermédiaire de servlets JSP. Cette architecture permet, d'une part, un accès banalisé à l'application sans installation à faire sur le poste client et, d'autre part, facilite le déploiement et la maintenance de l'application qui doit être uniquement installée sur un serveur.

CARAT communique via l'API JDBC avec trois bases de données relationnelles Postgres : la base de données de consommation contenant toutes les informations sur les consommations des individus, la base de données de contamination contenant les données de contamination des aliments et enfin, la base CARAT contenant les données prises en compte dans les calculs. CARAT utilise aussi pour les calculs statistiques l'API COLT développée par le CERN (Centre Européen pour la Recherche Nucléaire). L'API COLT propose une infrastructure Java pour le calcul scientifique. Elle contient entre autres des structures de données et des fonctionnalités utilisables pour l'analyse de données et les statistiques.

Les technologies que nous avons utilisées sont « open source » et gratuites.

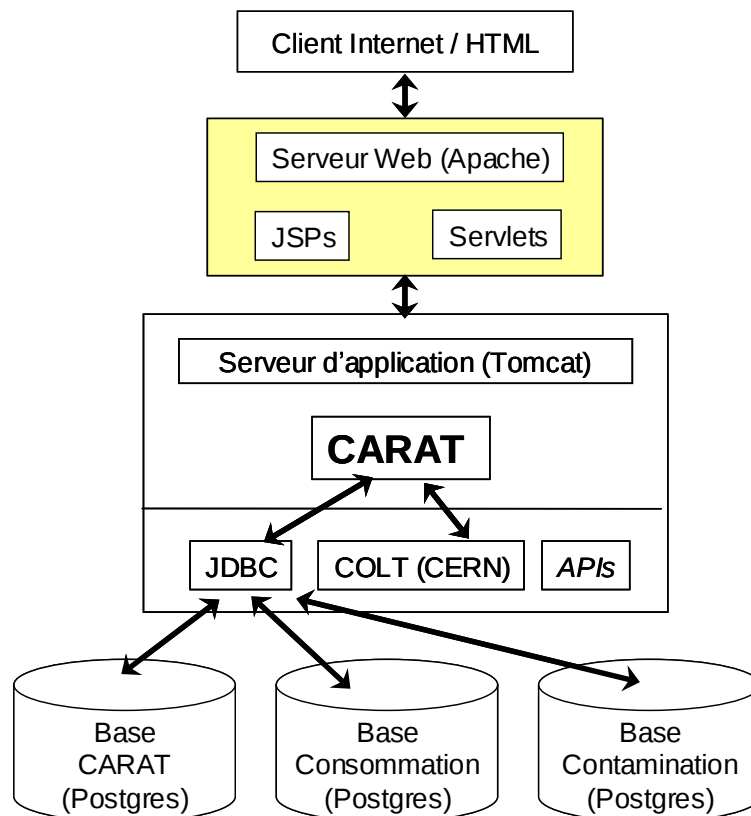


Figure 26. Architecture de CARAT

Dans sa forme actuelle, le logiciel CARAT permet d'évaluer les risques liés aux contaminants présents dans la Base de données de contamination de l'unité Mét@rik à partir de l'enquête individuelle de consommation INCA de 1999. Une première extension possible du logiciel consistera à permettre l'utilisation d'autres bases de données de consommation ou de composition des aliments, françaises ou étrangères. En ce qui concerne les données de consommation, l'intégration des données d'achats de ménage, comme le panel SECODIP, requiert le développement d'outils statistiques permettant l'individualisation des données de ménage. Par ailleurs, les nomenclatures des produits consommés et analysés sont dans la version actuelle du logiciel relativement proches mais le développement d'outils informatiques facilitant la comparaison de ces nomenclatures est crucial dès que l'on souhaite utiliser d'autres sources de données. Ces deux points font actuellement l'objet de recherches au sein de l'unité. L'individualisation des données de ménage est étudiée selon deux points de vue : l'utilisation de la structure d'âge et de sexe du ménage dans un modèle économétrique à partir des seules données de ménage ou bien l'apprentissage de la forme de la décomposition des expositions des ménages à partir de données d'expositions individuelles. Un module supplémentaire intégrant les outils relatifs à la théorie des valeurs extrêmes sera également développé. Ce module permettra d'évaluer de « petits risques » et de caractériser les populations les plus à risque pour éventuellement mettre en place des études plus ciblées sur ces populations et analyser l'impact de mesures sanitaires ou d'information vis à vis de ce risque. Ainsi, l'ensemble des méthodologies développées dans l'unité Mét@risk constitueront à terme des outils visant à améliorer les évaluations de risque ou de bénéfices (nutritionnels par exemple).

1.1.8. Bibliographie

Boer, W. J., H. van der Voet, P. E. Boon, G. Donkersgoed et J. D. Klaveren (2005). *MCRA a web-based program for Monte Carlo Risk Assessment. Manual Version 2005-04-26 Release 3.5. Technical report*. Biometris and RIKILT, Wageningen, The Netherlands.

Counil et al, 2005

Gibney M. J. and van der Voet H. (2003). *Introduction to the Monte Carlo project and the approach to the validation of probabilistic models of dietary exposure to selected food chemicals*, Food Additives & Contaminants, Volume 20, Supplement 1 / October 2003 Pages: 1 – 7.

Jessica Tressou, Méthodes Statistiques pour l'évaluation du risque alimentaire, thèse.