Risk Analysis, Vol. 22, No. 2, 2002

Quantitative Assessment of Exposure to the Mycotoxin Ochratoxin A in Food

Jean-Pierre Gauchi[1]* and Jean-Charles Leblanc[2]

This article presents the methodology and the simulation results concerning the quantitative assessment of exposure to the fungus toxin named Ochratoxin A (OA) in food, in humans in France. We show that is possible to provide reliable calculations of exposure to OA with the conjugate means of a nonparametric-type method of simulation, a parametric-type method of simulation, and the use of bootstrap confidence intervals. In the context of the Monte Carlo simulation, the nonparametric method takes into account the consumptions and the contaminations in the simulations only via the raw data whereas the parametric method depends on the random samplings from distribution functions fitted to consumption and contamination data. Our conclusions are based on eight types of food only. Nevertheless, they are meaningful due to the major importance of these foodstuffs in human nourishment in France. This methodology can be applied whatever the food contaminant (pesticides, other mycotoxins, Cadmium, etc.) when data are available.

KEY WORDS: Risk assessment; exposure; Ochratoxin A; food

1. INTRODUCTION

The aim of this study is to estimate, by Monte Carlo simulations, the exposure of the French population to Ochratoxin A (OA), a mycotoxin produced by fungi *Aspergillus Ochraceus* and *Penicillium Viridicatum*. This mycotoxin is a contaminant of grain stored in poor conditions, and through the food chain also contaminates several other foodstuffs, typically pork and poultry meat. Several toxicological studies[(1)] have shown the carcinogenic effect of this mycotoxin, particularly for the kidney, in rats. In humans, certain epidemiological studies indicate that this mycotoxin could be involved in the endemic nephropathy of the Balkans and cancer of the urinary tract, which is associated with it.[(2)] For these reasons, this mycotoxin remains under the close surveillance of the Conseil Supérieur d'Hygiène Publique in France. For more detailed toxicological information, the reader should consult the two references cited above. We add that the study presented here concerns only the static aspect of the exposure and not the dynamic aspect, which has to be conducted to evaluate a real risk level when OA-contaminated food is regularly ingested.

We will see that our aim was attained by the conjugate means of a nonparametric-type method of simulation (the NonParametric-NonParametric method, the *NP-NP* method), a parametric-type method of simulation (the Mixed Parametric-Parametric method, the *MP-P* method), and the use of bootstrap confidence intervals. If we consider the continuous probability density function (*pdf*) of the random variable ''normalized global exposure'' in a

[1] National Institute for Agricultural Research; Jouy-en-Josas; France Biometrics Unit; Food Risk Analysis Group.
[2] National Institute for Agricultural Research; Scientific Directorate for Human Nutrition and Food Safety.
* Address correspondence to Jean-Pierre Gauchi, INRA-Unité de Biométrie, Domaine de Vilvert, 78352 Jouy-en-Josas Cedex, France; telephone (33).1.34.65.28.47; gauchi@banian.jouy.inra.fr.

0272-4332/02/0400-0219$22.00/1

consumer population given, we have had to: (1) estimate the parameters of this *pdf* from the simulation outputs, a theoretical approach being untractable, (2) estimate several quantities such as mean, standard deviation, skewness, kurtosis, and quantiles, notably the 0.50th, 0.95th, and 0.99th quantiles, and their confidence intervals, from this fitted *pdf* or directly from the simulation outputs. The difficulty of our study was due to two causes: (1) the very irregular and asymmetrical raw data histograms and (2) we did not know the real contamination level of the ingested food during the survey week and we only had sparse contamination data (at our disposal) from various other studies in Europe. This second reason obliged us to carefully combine real data and simulation data to attain our goal.

Written from an essentially practical standpoint, this article is made up of the following sections. In Section 2 the raw data are described (consumptions and contaminations), Section 3 describes the exposure *NP-NP* and *MP-P* methods, and Section 4 provides the results of these methods. Section 5 compares the results concerning confidence interval methods and Section 6 is the conclusion. Simulations and calculations have been carried out by means of the Interactive Matrix Language (IML) and various procedures of the SAS software, Windows release 6.12 (SAS Institute®, Cary, NC, USA).

2. DESCRIPTION OF THE DATA

2.1. The Consumption Data

The *consumption data* of eight foodstuffs come from the ASPCC survey[3] conducted from June 1993 to June 1994 on 1,161 individuals. During seven days, the participants were questioned on their consumption of several types of food, which were weighed before eating. In Table I we describe the eight types of food studied here, keeping in mind that certain types cover several foods. For more information we refer the reader to Verger.[4]

Table I. Definition of the Eight Foodstuffs Considered in This Study

Food	Definition
CEREALS	rice, wheat, corn starch, bread, pasta, flour, crispbread, breakfast cereals, crackers, biscuits, pastries, croissants and the like, popcorn, cornflakes, cooked sweet corn
RAISINS	dried raisins
OTHER DRIED FRUIT (ODFRUIT)	apricots, bananas, dates, figs, prunes
PORK	kidneys, cutlets, fillet, roast, ribs, loin, black pudding, ham, pâté, salami, rillettes, and-ouillettes, bacon, sausages, foie gras
POULTRY	chicken liver, chicken, duck, goose
FRUIT JUICES (FRJUICES)	several kinds of fruit juices
WINES	several different wines
COFFEE	dry coffee (correction by a dilution coefficient of 17)

The sample of 1,161 individuals was divided into three classes: 377 men aged 15 or older, comprising 32.50% of the sample; 552 women aged 15 or older, comprising 47.50% of the sample; and 232 children younger than 15, comprising 20% of the sample. Detailed descriptive statistics and histograms can be found in Reference 5. For reasons of limited space, we only show some statistics (Table II) and the corresponding histogram (Fig. 1) of the children. The appearance of this latter histogram looks like the adult histograms. In Table II the percentage of consumers in relation to the individuals of the sample is also given. We see that very few individ-

Table II. Consumption in Grams Over an Average Day for the 232 children, with Percentage of Those Who Consume

	Food *N*	%	Mean	*SD*	Minimum	Maximum
CEREALS	232	100.00	195.36	89.87	26.14	619.00
RAISINS	17	7.33	3.39	3.00	0.43	11.71
ODFRUIT	15	6.46	6.29	5.12	1.29	17.14
PORK	219	94.40	41.44	24.87	0.57	140.00
POULTRY	167	71.98	22.76	15.68	1.43	95.71
FRJUICES	157	67.77	120.69	168.68	1.43	1497.14
WINES	25	10.77	4.94	6.00	0.43	23.86
COFFEE	24	10.34	3.33	3.36	0.14	12.57
BEER	6	2.59	25.88	27.93	2.14	71.43
VEGETABLES	75	32.33	15.36	9.74	1.00	47.14
SPICES	49	21.12	0.69	1.10	0.14	7.14

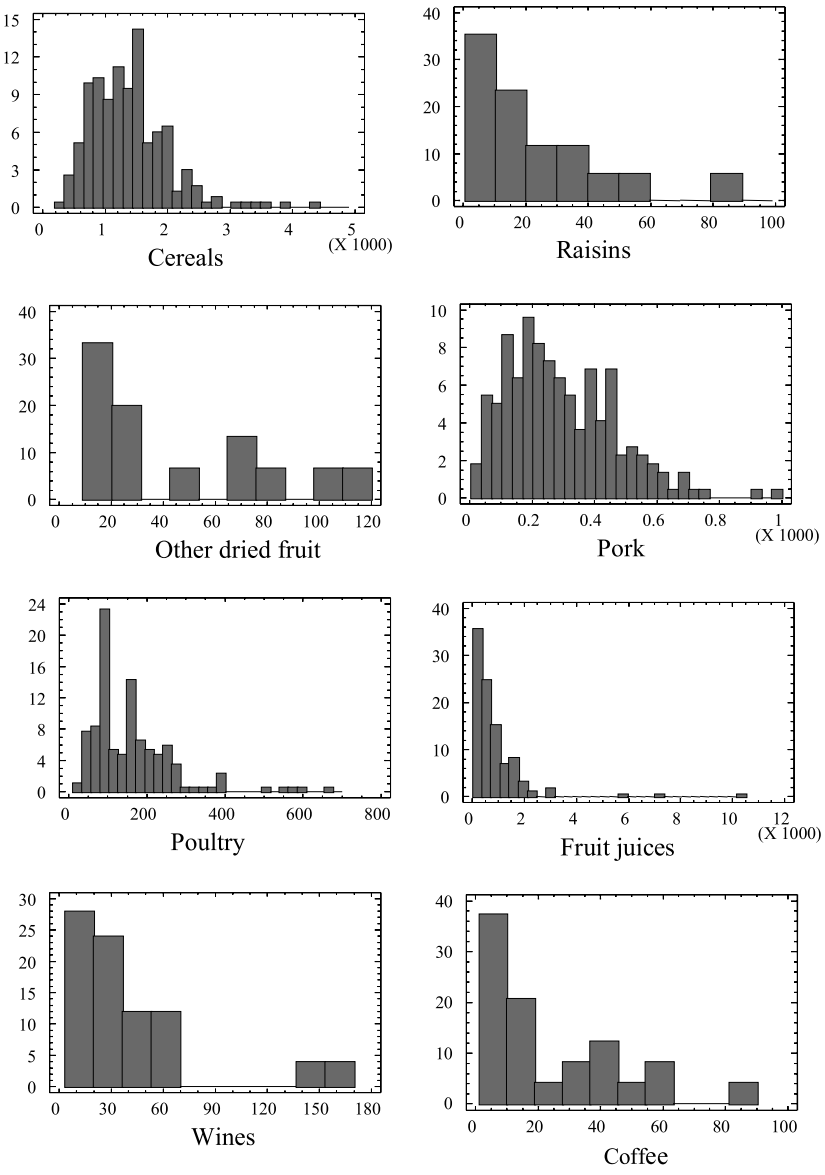

Fig. 1. Consumption histograms (relative frequencies) for the children who consume.

uals eat RAISINS and ODFRUIT, whereas everyone eats CEREALS. Considering the irregularity of the histograms, we propose to compare further on a nonparametric-type method of simulation (the *NP-NP* method) and a parametric-type method (the *MP-P* method).

As far as the consumption dependencies are concerned, scatter plots are useful for visualizing the patterns of one food consumption dependency versus another, after dividing the consumptions by the body weight in kilograms.[5] Since all possible scatter plots in this study showed only uniformly decreasing or increasing patterns, or no dependency at all, we could quantify the dependencies by Spearman correlation coefficients as more robust relative to the outliers than the Pearson correlation coefficients. The individuals who did not consume either of the two foods were excluded from the analysis. These correlations in the sense of Spearman were simulated by means of the Iman and Conover method[6] in the *MP-P* method (see Section 3.3). All the correlations have been retained, whether significant or not, to avoid adjusting a too arbitrary risk level. The corresponding coefficients for the three classes appear in Reference 5.

Table III. Descriptive Statistics of the Contamination Data of Each Food (in μg per kg of food)

Food	*N*	Mean	*SD*	Min	Max
CEREALS	183	0.65	0.98	0.20	7.20
RAISINS	13	0.71	1.16	0.10	4.30
ODFRUIT	33	0.28	0.29	0.10	1.60
PORK	1011	0.30	0.29	0.10	6.10
POULTRY	93	0.075	0.0342	0.00	0.18
FRJUICES	19	0.25	0.78	0.02	3.45
WINES	104	0.16	0.26	0.00	1.64
COFFEE	155	1.12	1.70	0.02	3.10

2.2. The Contamination Data

The *contamination data* used in our study come from elements independent of the previous consumption survey: they originate from SCOOP task 3.2.7;[7] see also Verger.[4] Table III gives the descriptive statistics of contamination and Fig. 2 shows the histograms of contamination for each of the eight types of food studied.

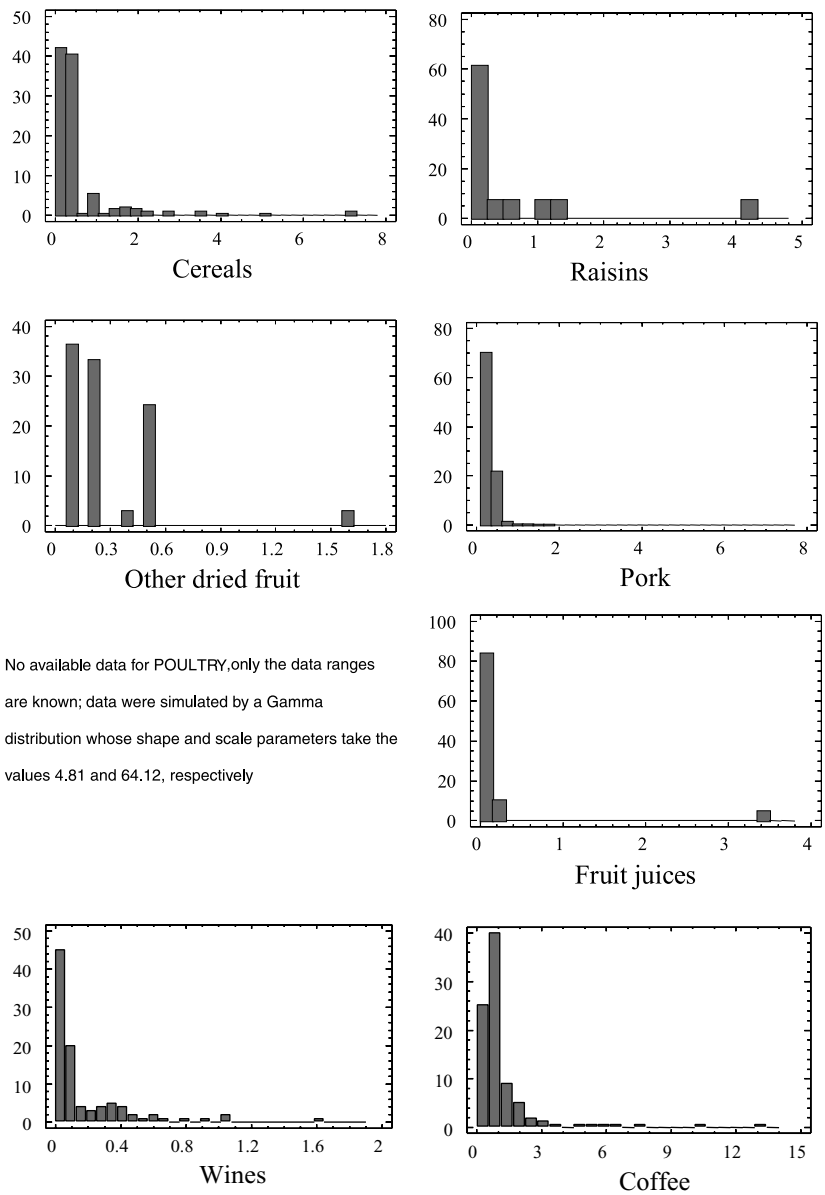

Fig. 2. Contamination histograms (relative frequences), in μg/kg for each food.

2.3. Remark on the Consumption and Contamination Histograms

The histograms shown in Fig. 1 and Fig. 2 are not optimal because their bar numbers are not determined according to a specific criterion. If more rigorous histograms are needed we advise using a mathematical criterion based on the Heillinger distance to construct them.[8–10] However, it is more useful to consider quantile-quantile plots (Q-Q plots) both to visualize distributions and to help in choosing the right distribution. We will see an example of such Q-Q plots in Section 3.3.2.

3. EXPOSURE CALCULATION METHODS

3.1. An Attempt at a Theoretical Approach

Relevant definitions are as follows:

$C_{\pi_*,j}$: The continuous random variable ''normalized consumption of the foodstuff j in a consumer population π_*.'' The term ''normalized'' means that the consumption is divided by the individual weight. Observations of this random variable are thus expressed in grams per kg of body weight per day, i.e., in g $\times$ bw$^{-1}\times$ day^{-1} units.

T_j: The continuous random variable ''OA level of the foodstuff j,'' respectively. Observations of this random variable are expressed in micrograms per kg of food, i.e., in µg $\times$ kg^{-1} units.

$E_{\pi_*}^{(NG)}$: The continuous random variable ''normalized global exposure in a consumer population π_*,'' the term ''global'' meaning that p foodstuffs are consumed. Observations of this random variable are expressed in nanograms per kg of body weight per day, i.e., in ng $\times$ bw$^{-1}\times$ day^{-1} units.

The random variable $E_{\pi_*}^{(NG)}$ is a function of the random variables $C_{\pi_*,j}$ and T_j as follows:

$$E_{\pi_*}^{(NG)} = \sum_{j=1}^{p} C_{\pi_*,j} T_j \tag{1}$$

In a theoretical approach we should be interested to know the closed form of the *pdf* of $E_{\pi_*}^{(NG)}$ as defined in Equation (1) and therefore be able to derive certain moments of the $E_{\pi_*}^{(NG)}$ distribution. Unfortunately Equation (1) is generally too complicated to reach this objective, notably if the *pdf* of the random variables $C_{\pi_*,j}$ and T_j are different and moreover not independent.

Finally, from a theoretical standpoint we can merely express and estimate the mean and variance of $E_{\pi_*}^{(NG)}$. Taking the mathematical expectation of Equation (1), whatever the dependency between the random variable products $C_{\pi_*,j}T_j$ and assuming independency between the consumptions $C_{\pi_*,j}$ and the contaminations T_j, we have:

$$\begin{aligned} \mathrm{E}(E_{\pi_*}^{(NG)}) &= \mathrm{E}\left(\sum_{j=1}^{p} C_{\pi_*,j} T_j\right) \\ &= \sum_{j=1}^{p} \mathrm{E}(C_{\pi_*,j})\mathrm{E}(T_j) \end{aligned} \tag{2}$$

and an unbiased estimation is given by:

$$\hat{\mathrm{E}}(E_{\pi_*}^{(NG)}) = \sum_{j=1}^{p} \hat{\mathrm{E}}(C_{\pi_*,j})\hat{\mathrm{E}}(T_j) \tag{3}$$

that is to say

$$\hat{E}_{\pi_0}^{(NG)^*} = \sum_{j=1}^{p} \bar{C}_{\pi_0,j}\bar{T}_j \tag{4}$$

where $\bar{C}_{\pi_0,j}$ is the arithmetical mean of the normalized consumptions of the food j in the sample π_0 (for example, the sample of the 232 children), and $\bar{T}_j$ is the arithmetical mean of the available contamination data of the food j.

If we consider that the consumption dependencies are low (as shown in Reference 5) and, moreover, if we assumed again no dependencies between the consumptions and contaminations (which is a reasonable hypothesis in this study), then we can express the variance of $E_{\pi_*}^{(NG)}$ by:

$$\mathrm{V}(E_{\pi_*}^{(NG)}) = \sum_{j=1}^{p} \mathrm{V}(C_{\pi_*,j}T_j) \tag{5}$$

where the variance of each product of two independent random variables is defined by the well-known formula:[11:261]

$$\begin{aligned} \mathrm{V}(C_{\pi_*,j}T_j) = \mathrm{V}(C_{\pi_*,j})\mathrm{V}(T_j) + \mathrm{E}(T_j)^2\mathrm{V}(C_{\pi_*,j}) \\ + \mathrm{E}(C_{\pi_*,j})^2\mathrm{V}(T_j) \end{aligned} \tag{6}$$

Replacing means and variances by their estimations calculated from the available data, the Equations (4) and (6) lead to the results given in Table IV.

In the general case, if we need estimations of more complicated parameters such as the quantiles, because of the above-mentioned intrinsic difficulties,

Table IV. Results of Exposure with the Theoretical Approach

Estimated Parameters	Children	Women	Men
$\hat{E}_{\pi_0}^{(NG)^*}$ (Equation (4))	5.849	3.040	3.498
$\hat{\mathbf{V}}(E_{\pi_0}^{(NG)})^*$ (Equation (6))	$(9.333)^2$	$(3.867)^2$	$(4.226)^2$

we will move toward the two following simulation methods. These methods will lead to the estimations of the theoretical *pdf* parameters as we will see later.

3.2. The *NP-NP* Method

3.2.1. Principle of the NP-NP Method

The *NP-NP* method we propose is a natural method of exposure assessment when consumption and contamination data are available. It is a completely nonparametric method in as much as no *pdf* hypothesis is made either on the consumption or on the contamination. Indeed, each normalized consumption profile of the survey is taken into account and each type of consumed food is attributed a value of contamination randomly drawn from the available contamination data.

3.2.2. Details of the NP-NP Method

Take:

- $c_{i(\pi_0),j}$: a daily normalized consumption $i(i = 1, \ldots, n_{\pi_0})$ of the foodstuff $j(j = 1, \ldots, p)$; it is a consumption made by an individual i of a certain weight, belonging to the sample π_0; the consumption is said to be normalized because it is divided by the individual weight,
- $t_{i(\pi_0),j}$: the real OA contamination level of the foodstuff j, consumed by an individual i of the sample π_0; as it is unknown it is therefore replaced by a random deviate $\tilde{t}_{i(\pi_0),j(F_j^{(o)})}$ drawn from the observed cumulative frequency of contamination $F_j^{(o)}$ of the foodstuff j; for each individual i, the p random deviates
- $\tilde{t}_{i(\pi_0),j(F_j^{(o)})}$ are independently sampled because, in this study, no dependency hypothesis is assumed between contaminations.

Then, an estimated normalized global exposure for p ingested food, with this *NP-NP* method, should thus be written:

$$\hat{E}_{i(\pi_0)}^{(NG)|[NP-NP]} = \sum_{j=1}^{p} c_{i(\pi_0),j} \tilde{t}_{i(\pi_0),j(F_j^{(o)})} \tag{7}$$

and the mean of $E_{\pi_*}^{(NG)}$ will be estimated by:

$$\hat{E}_{S(\pi_0)}^{(\mathrm{NG})|[\mathrm{NP-NP}]} = \frac{1}{n_{\pi_0}} \sum_{i=1}^{n_{\pi_0}} \frac{1}{n} \sum_{k=1}^{n} \times \left[\sum_{j=1}^{p} c_{i(\pi_0),j} \tilde{t}_{i(\pi_0),j^{(k)}(F_j^{(0)})} \right] \tag{8}$$

with:

- n_{π_0}, the number of individuals in the sample π_0,
- n, the number of contamination random deviates drawn from F_{0j} for each individual, here set at 1,000 (we checked that 10,000 random deviates do not improve the estimations),
- $\tilde{t}_{i(\pi_0),j^{(k)}(F_j^{(0)})}$, the k^{th} contamination random deviate drawn from $F_j^{(o)}$, for the individual i of the sample π_0,
- $S(\pi_0)$, the simulation set of the nn_{π_0} outputs calculated with Equation (8), noted only as S later, for simplification.

The $N(= nn_{\pi_0})$ outputs can lead to three types of estimation (or visualization) of the normalized global exposure *pdf*: An histogram or a Q-Q plot, a fitted Lognormal *pdf*, and a fitted Gamma *pdf*, the latter two chosen and fitted to the N simulation outputs by means of the methodology indicated in Section 3.3.2. Moreover, from the N outputs we calculated simulation statistics such as standard deviation, skewness and kurtosis coefficients, and exposure quantiles. From the fitted *pdf*s we obtained the equivalent estimated statistics and then we could compare them to the simulation statistics. We refer the reader to Kendall and Stuart[11:chap.2] for definitions of these usual statistics. Nevertheless, we give hereafter the definition of theoretical and empirical quantiles. If $F_{E_{\pi_*}^{(NG)}}$ stands for the theoretical *cdf* and $Q_{\pi_*}^{(NG),\alpha}$ for the α^{th} theoretical exposure quantile, this latter is here defined by: $F_{E_{\pi_*}^{(NG)}}(Q_{\pi_*}^{(NG),\alpha}) = \alpha$, whereas the α^{th} empirical exposure quantile $\hat{Q}_S^{(NG)|[NP-NP],\alpha}$ is defined by:

$$\#\left[\hat{E}_{i(S)}^{(NG)|[NP-NP]} \leq \hat{Q}_S^{(NG)|[NP-NP],\alpha}\right]/N = \alpha, \tag{9}$$

where the notation $\#[x_i \leq K]$ means the number of x_i less than or equal to K. For example, if $N = 1,000$, the estimate of the 0.95th quantile is the 950th largest value of all the $\hat{E}_{i(\pi_0)}^{(NG)|[NP-NP]}$. If αN is not an integer, the following procedure can be used. Assuming $\alpha \leq 0.5$, let $k = [(N+1)\alpha]$, the largest integer $\leq (N+1)\alpha$. Then we define the empirical α^{th} and $(1-\alpha)^{\text{th}}$ quantiles by the k^{th} largest and

$(N+1-k)^{\text{th}}$ largest values of all the $\hat{E}_{i(\pi_0)}^{(NG)|[NP-NP]}$, respectively.

We can refer to estimated quantiles from the fitted Lognormal and Gamma *pdf* as $\hat{Q}_{LN}^{(NG)|[NP-NP],\alpha}$ and $\hat{Q}_{\Gamma}^{(NG)|[NP-NP],\alpha}$, respectively. The results for the *NP-NP* method are given in the first parts of Tables V and VII.

3.3. The *MP-P* Method

3.3.1. Principle of the MP-P Method

Unlike the previous method, the *MP-P* method is essentially parametric in the sense that a mixed *pdf* is fitted to each food consumption and a parametric *pdf* to each food contamination. We can question the advantage of proposing such a method as well as the *NP-NP* method. The major argument, which can be suggested by the histograms in Fig. 1 and Fig. 2, is that the *NP-NP* method can possibly lead to less reliable estimations, especially of the high 0.95th and 0.99th quantiles, because these histograms show several totally empty regions. Accurate knowledge of these high quantiles is indeed crucial for toxicologists because they could correspond to dangerous exposure levels in humans. Thus, we understand the need to have at least two methods, which validate each other, for a more reliable estimation of these high quantiles. Finally, it appeared *a posteriori* that this parametric approach was also necessary when the confidence intervals were calculated by the parametric bootstrap method, as we shall see later.

Table V. Simulation Results of Exposure to OTA by the *NP-NP* Method and the *MP-P* Method

Statistics of the Simulation Outputs	Children	Women	Men
***NP-NP* Method**			
N	232,000	552,000	377,000
Mean	5.90	3.09	3.57
Standard deviation	9.44	3.88	4.27
Skewness	6.19	4.62	4.42
Kurtosis	61.86	31.24	28.26
Median	3.21	1.94	2.27
0.90th quantile	11.97	6.15	7.17
0.95th quantile	20.50	9.73	10.82
0.99th quantile	49.21	20.96	23.22
***MP-P* Method**			
N	100,000	100,000	100,000
Mean	5.94	3.10	3.57
Standard deviation	9.28	3.97	4.36
Skewness	4.65	3.93	3.57
Kurtosis	40.43	28.93	23.17
Median	2.67	1.77	2.13
0.90th quantile	14.99	7.27	8.31
0.95th quantile	22.35	10.33	11.68
0.99th quantile	44.43	19.32	21.11

3.3.2. Fitting of Probability Density Functions

We aim to fit the best *pdf*, among several possible candidate *pdfs*, on consumption and contamination data, for each sample π_0 (children, women, men). Regarding the very asymmetrical appearance of the histograms, we move toward the five following candidate distributions: Beta, Weibull, Exponential, Lognormal, and Gamma. To decide the best distribution among these five, for each of the consumption and contamination foodstuff data, our methodology was based on four steps:

1. For each of the five distributions a fitting was undertaken numerically with the method of maximum likelihood (using the CAPABILITY procedure of SAS software).
2. A quantile-quantile plot of each candidate distribution was exhibited. In the food context we can see, for instance, the paper of Smout *et al.*,[12] which investigates this point for normal distributions.

Table VI. Parameters of the Gamma *pdf* Fitted to the Subclasses of Normalized Consumptions

Foodstuff	Children ($\hat{r}$, $\hat{\lambda}$, $\hat{\theta}$)	Women ($\hat{r}$, $\hat{\lambda}$, $\hat{\theta}$)	Men ($\hat{r}$, $\hat{\lambda}$, $\hat{\theta}$)
CEREALS (0.439; 0.674)	2.594; 0.480; 1.610	4.151; 1.451; 0.286	3.901; 1.323; 0.570
RAISINS (0.375; 0.530)	0.701; 5.484; 0.010	0.589; 5.938; 0.004	0.567; 11.85; 0.008
ODFRUIT (0.998; 3.503)	0.582; 2.286; 0.043	0.673; 2.945; 0.013	0.538; 2.775; 0.024
PORK (1.085; 3.591)	1.714; 1.127; 0.041	2.041; 2.369; 0.046	2.118; 2.243; 0.031
POULTRY (4.810; 64.12)	1.813; 2.499; 0.080	1.955; 3.902; 0.044	1.772; 3.745; 0.031
FRJUICES (0.108; 0.423)	0.939; 0.209; 0.083	1.112; 0.629; 0.045	1.047; 0.734; 0.078
WINES (0.413; 2.404)	0.764; 5.594; 0.092	0.860; 0.415; 0.005	1.133; 0.307; 0.005
COFFEE (0.431; 0.386)	0.784; 8.961; 0.033	1.324; 5.240; 0.003	1.379; 6.018; 0.005

Notes: See the right-hand part of Fig. 5. In the first column, between parentheses appear $\hat{r}$ and $\hat{\lambda}$ of *pdf* fitted to the contamination data ($\hat{\theta} = 0$).

3. The Anderson-Darling statistic[13] was computed, taking a 10% (or less) first level risk to not accept the candidate distribution.
4. The agreement between empirical and estimated 0.90th, 0.95th, and 0.99th quantiles was considered.

Essentially, only the Gamma and Lognormal distributions have been retained in this methodology. Depending on the data, the Lognormal distribution was sometimes closer to the data than the Gamma distribution and sometimes it was the contrary, as we can see, for instance, on the Q-Q plots of Fig. 3. Because several presentations exist, we give hereafter the two closed forms of the Lognormal and Gamma *pdf*.

Suppose two continuous random variables X and Y linked by $X = \mathrm{Log}Y$. The Lognormal *pdf* for the random variable Y defined in $[0, +\infty]$ is:

$$g(y) = \frac{1}{\sqrt{2\pi} y \sigma_X} \exp\left(-\frac{1}{2}\left(\frac{Log(y) - m_X}{\sigma_X}\right)^2\right)$$

where m_X and σ_X are λ and r, the scale and shape parameters, respectively. Then X follows a Normal distribution with mean m_X and variance σ_X^2. The mean and the variance of Y are defined by $E(Y) = \exp(m_X + \sigma_X^2/2)$ and $V(Y) = [\exp(2m_X + \sigma_X^2)][\exp(\sigma_X^2) - 1]$, respectively.

The Gamma *pdf* for a continuous random variable X defined in $[0, +\infty]$ is:

$$f(x, r, \lambda) = \frac{1}{\lambda \Gamma(r)} \left(\frac{x - \theta}{\lambda}\right)^{r-1} \exp\left(-\frac{x - \theta}{\lambda}\right)$$

where r, λ, θ are the shape, scale, and threshold parameters, respectively, and $\Gamma(r)$ is the usual Euler's integral. The mean and variance of the Gamma distribution are connected to r and λ in the following way: $E(X) = r\lambda$ and $V(X) = r\lambda^2$.

3.3.2.1. Contaminations After the above methodology was applied, the fittings were chosen with Gamma densities whose estimated shape ($\hat{r}$) and scale ($\hat{\lambda}$) parameters are given in the first column of Table VI.

3.3.2.2. Consumptions CEREALS excepted, Table II shows that an often large proportion of individuals do not consume a given food. If we exaggerate, the typical histogram that we could plot for each foodstuff is shown in Fig. 4, where two zones appear corresponding to the two subclasses: nonconsumers (subclass of proportion h, represented by the tall bar on the left) and consumers (subclass of proportion $(1-h)$, represented by the right asymmetrical part). These proportions h are given in Table II (expressed in percentages) for each type of food in the children class.

In this situation we propose, for each sample π_0 and for each foodstuff j consumed, fitting a mixed distribution defined by:

$$U_{\pi_0,j}^{(D)} = \{[U(0, c_{i_{\min(\pi_0),j}})]j, h; [\Gamma(r, \lambda, \theta)]_{\pi_0,j}, (1 - h)\} \tag{10}$$

where:

- $[U(0, c_{i_{\min(\pi_0),j}})]_j$ is the continuous uniform distribution defined on the interval $[0, c_{i_{\min(\pi_0),j}}]$ with $c_{i_{\min(\pi_0),j}}$ the minimal consumption of the foodstuff j, in the sample π_0 (see Table II). This distribution corresponds to the left part of Fig. 5.
- $[\Gamma(r, \lambda, \theta)]_{\pi_0,j}$ is the Gamma distribution for the foodstuff j. The values for the estimated parameters are given in Table VI. This distribution corresponds to the right part of Fig. 5.
- $U_{\pi_0,j}^{(D)}$ means a sampling from a discrete uniform distribution: a random number u is drawn from a continuous uniform distribution defined on [0; 1]. If u is less than or equal to h then a new random number u' is drawn from $[U(0, c_{i_{\min(\pi_0),j}})]_j$, otherwise a new random number is drawn from $[\Gamma(r, \lambda, \theta)]_{\pi_0,j}$.

3.3.3. Details of the MP-P Method

Take:

- $\tilde{c}_{k(\hat{F}_{U,j}),j}$: a random normalized consumption for the foodstuff j; the random deviate k is drawn from $\hat{F}_{U,j}$ the corresponding *cdf* of the distribution consumption defined by Equation (10),
- $\tilde{t}_{k(\hat{F}_{T_j}),j}$: a random contamination for the foodstuff j; the random deviate k is drawn from $\hat{F}_{T_j}$, the fitted Gamma *cdf*, whose density parameters are given in the first column of Table VI.

Then, the estimated normalized global exposure k of the simulation set S should thus be written:

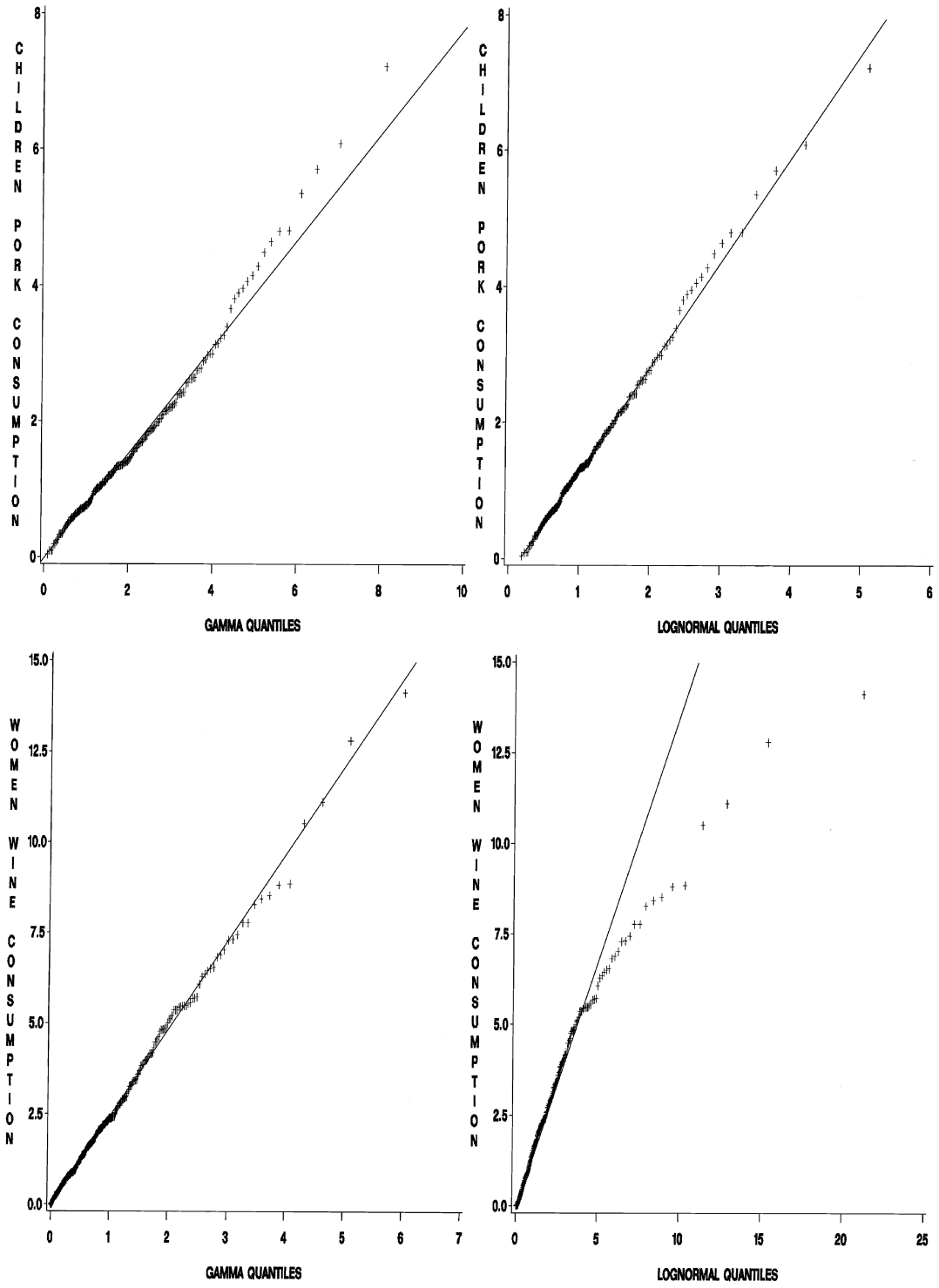

Fig. 3. Examples of Q-Q plots of Lognormal and Gamma distributions for consumption data.

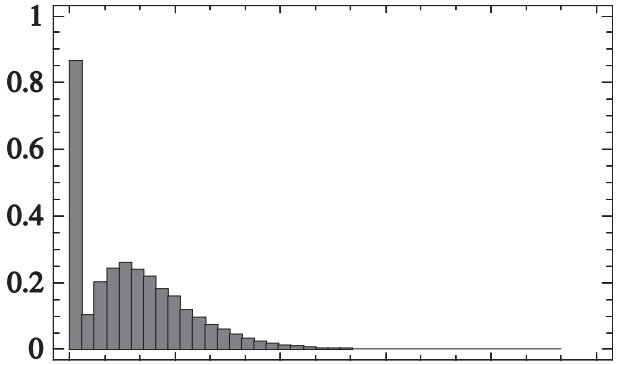

Fig. 4. Typical consumption histogram.

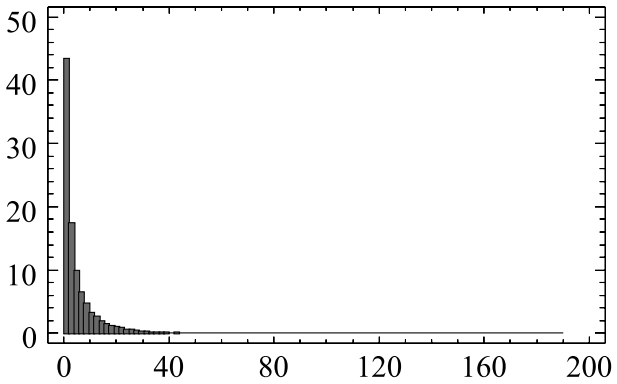

Fig. 5. Exposure output histogram (relative frequencies) with the *NP-NP* method, for the children.

$$\hat{E}_{k(S)}^{(NG)|[MP-P]} = \sum_{j=1}^{p} \tilde{c}_{k(\hat{F}_{U\,j}),j} \tilde{t}_{k(\hat{F}_{T_j}),j} \qquad (11)$$

and the mean of the normalized global exposures over the simulation set S will be estimated by:

$$\hat{E}_{S}^{(NG)|[MP-P]} = \frac{1}{n} \sum_{k=1}^{n} \hat{E}_{k(S)}^{(NG)|[MP-P]} \qquad (12)$$

with n as the number of random deviates drawn, here set at 100,000 (we checked that one million random deviates does not improve the estimations).

In the first step, the eight vectors of the n consumption random deviates were independently sampled. The second time they were arranged, by means of the Iman and Conover method,[6] to be correlated as in the correlation matrix given in Reference 5 in order to respect the consumption dependencies (see Section 2.1).

The results for the *MP-P* method are given in the second parts of Tables V and VII.

4. RESULTS OF THE EXPOSURE POINT ESTIMATIONS

4.1. Results of the Theoretical Approach

The results of the theoretical approach are given in Table IV.

4.2. Results of the *NP-NP* Method

The results of the *NP-NP* method appear in the first part of Table V. On one hand, we observe the very large skewness and kurtosis coefficients and on the other hand the large figure for the 0.99th quantile of the children class. We show in Fig. 5 the histogram of the simulation outputs for the children class. In the first part of Table VII appear Lognormal and Gamma *pdf* fitted to the simulation outputs, and the corresponding statistics.

We checked the negligible influence of the random generator seed on the main results: for example, with 10 different seeds $\hat{E}_{children}^{(NG)|[NP-NP]}$ varies in the interval [5.78; 5.97], $\hat{E}_{women}^{(NG)|[NP-NP]}$ varies in the interval [2.97; 3.08] and $\hat{E}_{men}^{(NG)|[NP-NP]}$ varies in the interval [3.41; 3.53]. Thus, the *NP-NP* method leads to estimations of $\mathrm{E}(E_{\pi_*}^{(NG)})$, which are only very slightly biased because we can see that the values of $\hat{E}_{\pi_0}^{(NG)^*}$, given in Table IV, fall into this latter interval.

Other results concerning the confidence intervals for the *NP-NP* method will be given in Section 5. The Anderson-Darling statistic shows that a Lognormal distribution hypothesis is rejected whereas the Gamma distribution hypothesis is accepted. However, the estimated statistics are close to each other for both hypotheses. For children, the empirical values of Table V are poorly representative; for example, compare the values of standard deviation and 0.99th quantile of Tables V and VII.

4.3. Results of the *MP-P* Method

The parameters of the Gamma *pdf* fitted to consumption and contamination data are given in Table VI. The results of the *MP-P* method are given in the second parts of Tables V and VII. The histograms of the simulation outputs are not given here but they look like those obtained with the *NP-NP* method (see Fig. 5).

Table VII. Estimated Statistics of Exposure to OTA by Lognormal and Gamma *pdf* Fitted to the Outputs of the *NP-NP* Method and the *MP-P* Method

Estimated Statistics	Children	Women	Men
***NP-NP* Method**			
LOGNORMAL DISTRIBUTION	$\hat{r} = 0.871$; $\hat{\lambda} = 1.286$	$\hat{r} = 0.772$; $\hat{\lambda} = 0.758$	$\hat{r} = 0.755$; $\hat{\lambda} = 0.924$
Estimated mean	5.29	2.87	3.35
Estimated *SD*	5.64	2.59	2.93
Estimated median	3.62	2.13	2.52
Estimated 0.90th quantile	11.06	5.74	6.62
Estimated 0.95th quantile	15.17	7.60	8.71
Estimated 0.99th quantile	27.48	12.86	14.57
AD (*p*-value)	2419 (< 0.005)	5620 (< 0.005)	3553 (< 0.005)
GAMMA DISTRIBUTION	$\hat{r} = 1.160$; $\hat{\lambda} = 0.196$	$\hat{r} = 1.509$; $\hat{\lambda} = 0.489$	$\hat{r} = 1.581$; $\hat{\lambda} = 0.443$
Estimated mean	5.94	3.08	3.57
Estimated *SD*	5.48	2.51	2.84
Estimated median	4.32	2.43	2.85
Estimated 0.90th quantile	13.10	6.42	7.34
Estimated 0.95th quantile	16.78	8.03	9.13
Estimated 0.99th quantile	25.25	11.65	13.17
AD (*p*-value)	10533 ($\geq$ 0.10)	22273 ($\geq$ 0.10)	13497 ($\geq$ 0.10)
***MP-P* Method**			
LOGNORMAL DISTRIBUTION	$\hat{r} = 1.421$; $\hat{\lambda} = 0.914$	$\hat{r} = 1.142$; $\hat{\lambda} = 0.534$	$\hat{r} = 1.116$; $\hat{\lambda} = 0.709$
Estimated mean	6.85	3.27	3.79
Estimated *SD*	17.50	5.36	5.95
Estimated Median	2.49	1.71	2.03
Estimated 0.90th quantile	15.41	7.37	8.49
Estimated 0.95th quantile	25.81	11.15	12.74
Estimated 0.99th quantile	68.02	24.29	27.24
AD statistics (*p*-value)	125 (< 0.01)	52 (< 0.005)	71 (< 0.005)
GAMMA DISTRIBUTION	$\hat{r} = 0.644$; $\hat{\lambda} = 0.108$	$\hat{r} = 0.971$; $\hat{\lambda} = 0.314$	$\hat{r} = 1.020$; $\hat{\lambda} = 0.285$
Estimated mean	5.94	3.09	3.57
Estimated *SD*	7.11	3.14	3.54
Estimated median	3.45	2.12	2.50
Estimated 0.90th quantile	14.92	7.18	8.19
Estimated 0.95th quantile	20.24	9.37	10.63
Estimated 0.99th quantile	32.95	14.47	16.30
AD statistics (*p*-value)	940 ($\geq$ 0.10)	812 ($\geq$ 0.10)	680 ($\geq$ 0.10)

Note: AD stands for Anderson-Darling statistics.

An important comment on Table V is to outline that the estimation of $E_{\pi_*}^{(NG)}$ by means of the *MP-P* method is very close to the estimation obtained with the *NP-NP* method and also close to those of Table IV, the small difference coming from the usual random generator variability of a simulation study. This observation indirectly confirms the validity of the *pdf* fitted to consumption and contamination data. The estimations of the high quantiles with the *MP-P* method are close to those obtained with the previous method but it is necessary to provide confidence intervals to check if it is true or not. This will be done in the next section. The main difference between the two methods in Table V concerns the estimation of skewness and kurtosis coefficients for children.

Another comment can be made about the consumption dependencies in the simulations. Taking these dependencies into account or not did not significantly change the results. For example, the children in the 0.95th and 0.99th quantiles, which have the values 22.35 and 44.43, respectively, when the dependencies are not taken into account, have the values 23.14 and 47.31 when the dependencies are taken into account. This behavior has several origins: first, the correlation coefficients are small (as shown in Reference 5); second, we assumed independent contaminations; and finally we calculated exposures from the products of consumption random deviates with contamination random deviates. Therefore, the low dependency between the consumptions is somewhat lost in a predominant independency.

The second part of Table VII shows Lognormal and Gamma *pdf* fittings to the simulation outputs, and the corresponding statistics.

In Table VII it appears that the estimated statistics are quite different between the *MP-P* method and the *NP-NP* method. For example: for the children class, the values 6.85, 17.50, and 68.02 for the mean, standard deviation, and 0.99th quantile, respectively, obtained with the Lognormal distribution, seem to be inflated in the *MP-P* method. In this case the use of the Lognormal fitting is totally inadequate, whereas the Gamma fitting is quite well adapted as we can see in Table VII. On the other hand, the values of 14.47 and 16.30 for the estimation of 0.99th quantiles for women and men, respectively, obtained with the Gamma distribution, are very different from those of Table V, whereas, this time, the Lognormal seems more adequate with the *MP-P* method.

It is well known that a reliable estimation of high quantiles is very difficult when the data show histograms with very large skewness and kurtosis coefficients as is the case here. However, a lot of work has been done in this direction. For instance, Reference 14 says that it is difficult or even impossible to estimate high quantiles without restricting to (at least) a semiparametric model. As pointed out by a referee, with such weak restrictions, methods are available.[15,16] We therefore recommend comparing the methods proposed in References 15 and 16 with our methods.

With these two *NP-NP* and *MP-P* methods, we now have some exposure point estimations at our disposal but we do not yet know their validity and, in particular, we do not have information about their accuracy. The aim of the next section is to calculate confidence intervals for the parameters that have been estimated by these statistics.

It should be noted that we checked that the use of the importance sampling method[17,18] for calculating of the statistics did not improve the results. The reason probably depends on the large number of iterations used for each statistic (that is to say 100,000) and also on the too poor distribution fittings in the extreme tail.

5. CONFIDENCE INTERVALS

The objective of this section is to provide confidence intervals for the parameters $\mathrm{E}(E_{\pi_*}^{(NG)})$, $\mathrm{V}(E_{\pi_*}^{(NG)})$, $Q_{\pi_*}^{(NG),0.50^{\text{th}}}$, $Q_{\pi_*}^{(NG),0.95^{\text{th}}}$, and $Q_{\pi_*}^{(NG),0.99^{\text{th}}}$. Two approaches have been considered to build confidence intervals for these parameters: a parametric bootstrap approach and, only for the $Q_{\pi_*}^{(NG),0.95^{\text{th}}}$ and $Q_{\pi_*}^{(NG),0.99^{\text{th}}}$ quantiles, an analytical approach.

5.1. An Attempt of Nonparametric Bootstrap Confidence Intervals

Before giving details on the parametric bootstrap approach we chose, we must explain why the classical nonparametric bootstrap approach is not possible here. No theoretical aspects on the bootstrap will be presented here; we refer the reader to Efron and Tibshirani[19] and Hall.[20] However, we very briefly recall the principle of the classical nonparametric bootstrap. Suppose we have a sample of n data. A bootstrap sample is obtained by randomly sampling n times, with replacement, from the n original data. A large number B of independent bootstrap samples is generated this way, each of them being of size n. For each bootstrap sample, a bootstrap replication of a given statistic s is obtained, for example, a standard error of a mean. With the B values of s we have at our disposal a bootstrap distribution of s that enables us to calculate empirical nonparametric bootstrap confidence intervals of s, for example, a confidence interval for a standard error.

This nonparametric bootstrap approach is rather complicated to apply here because after the consumption data samples of each consumer sample π_* and the eight contamination samples have been bootstrapped, it is necessary to draw randomly n (for instance $n = 1,000$) deviates from contamination data in order to make a simulation set. This simulation set enables, for example, the calculation of empirical quantiles (by means of Equation (9)). This procedure must be repeated B times and then the B quantile values make a bootstrap distribution. We observe that this procedure should not be a genuine nonparametric bootstrap because the n random drawings are necessary at each step. Indeed, we checked that this supplementary (but necessary) drawing resulted in very biased statistics, even for the mean, and therefore makes this hybrid nonparametric bootstrap approach unacceptable. Thus, we moved toward a parametric (or a pseudo-parametric) approach.

5.2. Parametric Bootstrap Confidence Intervals

In the parametric bootstrap, the B samples of size n are drawn from a *cdf* fitted to the original data sample. So, ''original'' data are now the exposure simulation outputs of the *NP-NP* or *MP-P* method. Because these data are not real data, but only simulation outputs, the bootstrap approach cannot be nonparametric but is instead parametric. We used three types of parametric bootstrap confidence intervals: a pseudo-parametric bootstrap confidence interval (defined hereafter as type 1), a Lognormal parametric bootstrap confidence interval (type 2), and a Gamma parametric bootstrap confidence interval (type 3). These three types are defined below.

- *Type 1: A pseudo-parametric bootstrap confidence interval*. We randomly draw B samples of size n_{π_0} in the exposure simulation set S (of the *NP-NP* or the *MP-P* method, separately), typically B equals 10,000. The size of S is for example 232,000 for the children class with the *NP-NP* method and 100,000 for the three classes with the *MP-P* method. It is called a pseudo-parametric bootstrap because no distribution is fitted to the S outputs, but, because of the huge size of S, it looks like a drawing by an algorithm for a specified theoretical distribution.
- *Type 2: A Lognormal parametric bootstrap confidence interval*. To the sets S are fitted Lognormal distributions whose estimated parameters are given in Table VII. Then, B samples of size n_{π_0} are randomly drawn from these fitted distributions. Typically B equals 10,000.
- *Type 3: A Gamma parametric bootstrap confidence interval*. It is equivalent to type 2, except this is the Gamma distribution in place of the Lognormal.

For each type the bounds of a 95%-confidence interval are calculated (with Equation (9)) taking the 0.025th and 0.975th empirical percentiles of the final bootstrap distribution. The results are given in Tables VIII and IX.

5.3. Analytical Confidence Intervals

In Reference 12 the authors construct one-sided intervals for the 0.95th quantile of their parameters studied by an analytical parametric procedure (Reference 12:688) and an analytical nonparametric procedure obtained from Shorack and Wellner.[21] Their empirical distributions were roughly normal. Because our own context is far from the normal case, we propose adapting to our problem two other types of analytical confidence intervals, described in detail elsewhere. Hereafter, we call them types 4 and 5.

Table VIII. Confidence Intervals Obtained from the Results of the *NP-NP* Method

95%-Confidence Interval	Children	Women	Men
Mean			
Type 1	[4.83; 7.20]	[2.77; 3.42]	[3.16; 4.02]
Type 2	[4.62; 6.06]	[2.66; 3.10]	[3.06; 3.66]
Type 3	[5.23; 6.63]	[2.87; 3.29]	[3.28; 3.86]
Standard deviation			
Type 1	[5.67; 14.63]	[2.99; 4.83]	[3.14; 5.50]
Type 2	[4.05; 8.07]	[2.14; 3.19]	[2.36; 3.71]
Type 3	[4.59; 6.47]	[2.26; 2.78]	[2.50; 3.19]
0.50th Quantile			
Type 1	[2.84; 3.62]	[1.79; 2.07]	[2.09; 2.47]
Type 2	[3.12; 4.14]	[1.96; 2.31]	[2.29; 2.77]
Type 3	[3.62; 5.08]	[2.21; 2.67]	[2.55; 3.18]
0.95th Quantile			
Type 1	[14.07; 29.56]	[7.84; 12.05]	[8.57; 14.08]
Type 2	[12.13; 19.45]	[6.66; 8.71]	[7.50; 10.33]
Type 3	[14.09; 20.11]	[7.23; 8.88]	[8.11; 10.37]
Type 4	[13.39; 29.49]	[7.78; 11.93]	[8.40; 13.65]
Type 5	[13.83; 19.73]	[7.20; 8.86]	[8.01; 10.25]
0.99th Quantile			
Type 1	[28.08; 76.92]	[15.35; 27.57]	[16.16; 33.70]
Type 2	[18.47; 40.81]	[10.21; 16.40]	[11.37; 20.14]
Type 3	[19.40; 32.16]	[10.02; 13.67]	[11.08; 16.28]
Type 4	[27.16; 81.95]	[14.97; 26.79]	[15.68; 30.65]
Type 5	[10.72; 39.78]	[7.60; 15.70]	[7.74; 18.60]

Note: See the text for the definition of the different types.

Type 4: A distribution-free analytical confidence interval. A distribution-free analytical confidence interval was proposed by Wilks.[22] The rather complicated formulas given for small samples in this reference are replaced by simpler formulas when the samples are large (Reference 22:Equations 11.2.9–11.2.13), which is the case here ($n_{\pi_0} \geq 232$). Then, an approximate 95%-confidence interval is given by:

$$\text{Prob}\left(\hat{Q}\left(\left[n_{\pi_0}\alpha_{LOW}^{0.95}\right]\right) < Q_s^{\alpha} < \hat{Q}\left(\left[n_{\pi_0}\alpha_{UP}^{0.95}\right]\right)\right) \cong 0.95 \tag{13}$$

where the two order statistics $\hat{Q}\left(\left[n_{\pi_0}\alpha_{LOW}^{0.95}\right]\right)$ and $\hat{Q}\left(\left[n_{\pi_0}\alpha_{UP}^{0.95}\right]\right)$ constitute the bounds of an approximate 95%-confidence interval for the quantile Q_s^{α}

Table IX. Confidence Intervals Obtained from the Results of the *MP-P* Method

95%-Confidence Interval	Children	Women	Men
Mean			
Type 1	[4.83; 7.20]	[2.77; 3.44]	[3.15; 4.02]
Type 2	[5.10; 9.47]	[2.86; 3.74]	[3.25; 4.44]
Type 3	[5.03; 6.93]	[2.84; 3.36]	[3.23; 3.95]
Standard deviation			
Type 1	[6.43; 13.65]	[3.22; 4.99]	[3.45; 5.57]
Type 2	[7.88; 34.49]	[3.68; 8.23]	[3.94; 9.23]
Type 3	[5.91; 9.06]	[2.77; 3.53]	[3.06; 4.06]
0.50th Quantile			
Type 1	[2.08; 3.38]	[1.56; 1.99]	[1.83; 2.46]
Type 2	[1.96; 3.11]	[1.51; 1.91]	[1.77; 2.35]
Type 3	[2.53; 4.13]	[1.87; 2.38]	[2.15; 2.88]
0.95th Quantile			
Type 1	[16.86; 30.28]	[8.77; 12.28]	[9.76; 14.40]
Type 2	[17.94; 38.75]	[9.17; 13.64]	[10.20; 16.37]
Type 3	[16.63; 25.85]	[8.26; 10.57]	[9.24; 12.41]
Type 4	[16.30; 29.75]	[7.65; 10.52]	[8.44; 12.40]
Type 5	[15.98; 24.50]	[8.22; 10.52]	[9.09; 12.17]
0.99th Quantile			
Type 1	[22.52; 66.29]	[15.02; 25.11]	[16.08; 29.62]
Type 2	[35.60; 129.6]	[17.27; 34.81]	[18.86; 43.90]
Type 3	[24.73; 45.73]	[12.08; 17.39]	[13.39; 20.80]
Type 4	[27.89; 72.10]	[12.59; 20.48]	[13.76; 23.78]
Type 5	[12.94; 52.95]	[8.68; 20.26]	[8.59; 24.00]

Note: See the text for the definition of the different types.

where $\left[n_{\pi_0}\alpha_{LOW}^{0.95}\right]$ and $\left[n_{\pi_0}\alpha_{UP}^{0.95}\right]$ are the largest integers in $n_{\pi_0}\alpha_{LOW}^{0.95}$ and $n_{\pi_0}\alpha_{UP}^{0.95}$, respectively, and $\alpha_{LOW}^{0.95}$ and $\alpha_{UP}^{0.95}$ probabilities, which can be determined as indicated in Reference 22.

We propose applying this principle to our problem by means of the following three steps:

1. Random sampling of m subsets of size $n_{\pi o}$ from the set S (originating from the *NP-NP* or *MP-P* method), assuming its *cdf* is an estimation of the true exposure continuous *cdf*. A typical value for m is 10,000.
2. For each of the m subsets calculating a 95%-confidence interval for the quantile of order α, with Equation (13).
3. Averaging the 10,000 lower and upper bounds, separately, of the previous intervals. Finally, we have an approximate 95%-confidence interval for the quantile $Q_{\pi_*}^{(NG),\alpha}$.

The results are given in Tables VIII and IX.

Type 5: A parametric analytical confidence interval. This second analytical confidence interval is a parametric confidence interval because its calculation is based on the parameters of the fitted density to the relative frequencies.[2,3,11:251–252] With the same notations as for type 4, we recall now the results. From the asymptotic distribution of an empirical quantile of order α, an asymptotic 95%-confidence interval for Q_s^α is given by

$$\hat{Q}_s^\alpha \pm 1.96\sqrt{\alpha(1-\alpha)/(n_s\hat{f}(\hat{Q}_s^\alpha)^2)} \qquad (14)$$

where $\hat{f}(\hat{Q}_s^\alpha)$ is the ordinate of the fitted *pdf* at the abscissa $\hat{Q}_s^\alpha$. To apply this principle to our problem, the same steps as for type 4 are used, except for step 2, which becomes: For each of the m subsets fitting a Gamma distribution and then calculating a 95%-confidence interval for the quantile of order α with Equation (14).

The results are given in Tables VIII and IX.

5.4. Results and Discussion of the Confidence Intervals

For Tables VIII and IX, it is obvious that the Lognormal and Gamma fittings are not good enough because the intervals of types 2, 3, and 5 do not always contain the corresponding estimated parameter. For example, in Table VIII, we can see that the type 3 confidence interval for the children standard deviation [4.59; 6.47] does not contain the value 9.44 (see Table V). Many other examples can be found in Tables VIII and IX, especially for the quantiles. On the contrary, types 1 and 4 always contain the values of the estimated parameters and, moreover, their lower and upper bounds are always close to each other, respectively. To be the closest to the simulation context, we decided that type 1 confidence interval was the best choice.

6. CONCLUSION

6.1. Methodological Aspects

From a statistical standpoint, the quantitative assessment of exposure to OA is not easy to achieve with raw data showing such irregular and asymmetrical histograms, especially if confidence intervals are required. Therefore, we propose, on one hand, two different methods, the *NP-NP* method and the *MP-P* method, and, on the other hand, pseudo-parametric bootstrap confidence intervals (type 1 as defined above), to estimate this exposure with good

Table X. Summary Results for Exposure to OTA: Mean, Standard Deviation, and Quantiles

Parameters	Children	Women	Men
Mean	[4.83←5.85→7.20]	[2.77←3.04→3.42]	[3.16←3.50→4.02]
SD	[5.67←9.44→14.63]	[2.99←3.88→4.83]	[3.14←4.27→5.50]
0.95th quantile	[14.07←20.50→29.56]	[7.84←9.73→12.05]	[8.57←10.82→14.08]
0.99th quantile	[28.08←49.21→76.92]	[15.35←20.96→27.57]	[16.16←23.22→33.70]

Notes: The central figure is the estimation of the parameter coming from the *NP-NP* method, and the bounds define a 95%-confidence interval of type 1 for the parameter, calculated with the outputs of the *NP-NP* method.

reliability. In the present situation the usual parameters (means, standard deviations, etc.) are estimated well by both methods, even though for the children class the calculation of a confidence interval for the 0.99th quantile still remains a difficulty. On theoretical aspects of extreme upper quantiles we refer the reader to the references given.[24–26]

Nevertheless, from a practical point of view we can propose the following methodology to estimate the characteristics of the exposure distribution of a given population, when such data types are available and no determination of a closed form of its *pdf* is possible:

1. Calculation of estimations of the mean and the variance with Equations (4) and (6). Generalizations of Equation (6) exist if correlations are not negligible.[11:261]
2. Calculation of mean, standard deviation, median, skewness, kurtosis, and the high quantiles required, from the outputs of the *NP-NP* and *MP-P* methods, in order to validate them to each other, if the results are close.
3. Calculation of pseudo-parametric bootstrap confidence intervals (type 1) for the parameters corresponding to the preceding statistics.

To better assess the stability of the high quantiles and their confidence intervals, a sensitivity analysis relative to the fitted *pdf* parameters, for the class of children, is in progress and will soon be published.

6.2. Toxicological Aspects

We summarize in Table X the figures that we propose for estimating the exposure to OA of French children, women, and men. With a 5% risk of being wrong, we can say that exposure to OA is roughly the same for women and men (their confidence intervals overlap) whereas children are more exposed.

We can compare these figures to those provided in the work cited in the Introduction,[1] where the mean exposure for adults was found to be $1.3\,\text{ng} \times \text{bw}^{-1} \times \text{day}^{-1}$ and a 95th quantile for children of $6.90\,\text{ng} \times \text{bw}^{-1} \times \text{day}^{-1}$. Our own figures are higher and are worrying, particularly in regards to children. According to the authors, as far as the nephrotoxic effect is concerned, the Daily Tolerable Dose (DTD) is $14\,\text{ng} \times \text{bw}^{-1} \times \text{day}^{-1}$, and if we consider the carcinogenic effects, we obtain a Virtually Sure Dose of $1.8\,\text{ng} \times \text{bw}^{-1} \times \text{day}^{-1}$. However we outline that this DTD level is not an exposure as we calculated in our study and a strict comparison cannot be made between our figures and this DTD level. Our figures cannot be considered as long-term exposure because we have just taken into account the data over a weekly survey and no other weeks in the year have been studied here. Nevertheless, considering these threshold values and the high toxicity of OA, as well as the already high exposure figures observed, it seems to us crucial that other teams, particularly in other countries, compare their results with our own figures.

In this article we have deliberately given priority to the practical standpoint for providing a useful methodology to the decisionmaker. Obviously, some research studies have to be undertaken on the theoretical aspects mentioned in the quoted references. Moreover, we have to emphasize that our static approach is not sufficient, but it is the first step toward the objective of having a full risk method. A dynamic approach should be necessary to model the absorption and releasing phenomena of OA in humans. A type of dynamic approach is undertaken by another researcher.[27]

ACKNOWLEDGMENTS

We thank J. B. Denis, A. Kobilinsky, and I. Albert (INRA, Jouy-en-Josas) and B. Goffinet (INRA, Toulouse) for their precious advice in

improving the first version of this document. We also thank P. Verger (INRA, Paris).

Finally, I wish to thank the referees for their helpful comments, which led to an improved version of this article, especially one referee who advised me to the use of Q-Q plots and indicated very useful references.

REFERENCES

1. Pfohl-Leszkowicz, A. (1999). *Les mycotoxines dans l'alimentation: évaluation et gestion du risque*. TECDOC, Paris.
2. Moll, N., & Moll., M. (1995). *Sécurité alimentaire du consommateur*. TECDOC, Paris.
3. ASPCC. (1999). *Observatoire des Consommations Alimentaire: Etude des distributions statistiques des consommations de huit groupes d'aliments susceptibles de contenir de l'Ochratoxine A*. Internal Technical Report of CREDOC, for INRA, based on data from ASPCC Survey.
4. Verger, P. (1999). Exposure assessment to certain contaminants, working group ''contaminants.'' *Internal Technical Report*, March 1999. INRA, Paris.
5. Gauchi, J. P. (2000). Evaluation quantitative de l'exposition à l'Ochratoxine A. *Internal Technical Report*, corrected and enlarged 2nd ed., 4 September 2000. INRA INRA/Jouy-en-Josas/Biometrics Unit.
6. Iman, R. L., & Conover, W. J. (1982). A distribution-free approch to inducing rank correlation among input variables. *Commun. Statist.-Simula. Computa.*, **11** (3), 311–334.
7. SCOOP Task 3.2.7. Report of the European Community, French Contribution. (In press). Origin of data: Direction Générale de l'Alimentation (Ministère de l'Agriculture et de la Pêche), Direction Générale de la Concurrence, de la Consommation et de la Répression des Fraudes (Ministère de l'Economie, des Finances et de l'Industrie).
8. Donoho, D., Johnstone, I., Kerkyacharian, G., & Picard, D. (1996). Density estimation by wavelet thresholding, *Ann. Stat.*, **24** (2), 508–539.
9. Castellan, G. (1999). *Modified Akaike's criterion for histogram density estimation*. Preliminary report no. 61 of Orsay University, France.
10. Birgé, L., & Rozenholc, Y. *How many bins should be put in a regular histogram?* To be submitted.
11. Kendall, M., & Stuart, A. (1977). *The advanced theory of statistics*. Volume 1, 4th ed. Charles Griffin & Co. Ltd.
12. Smout, C., Van Loey, A., Hendrickx, M., & Beirlant, J. (2000). Statistical variability of heat penetration parameters in relation to process design. *Journal of Food Science*, **65** (4), 685–693.
13. D'Agostino, R. B., & Stephens, M. A. (1986). *Goodness-of-fit techniques*. Marcel Dekker, Inc. New York and Basel. pp. 100–101.
14. Beirlant, J., & Devroye, L. (1999). On the impossibility of estimating densities in the extreme tail. *Statist. Proba. Lett.*, **43**, 57–64.
15. Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events: For insurance and finance*. Springer Verlag.
16. Beirlant, J., Teugels, J. L., & Vynckier, P. (1996). *Practical analysis of extreme values*. Leuven University Press.
17. Clark, C. E. Importance sampling in Monte Carlo analysis. *Operational Research*, **9**, 603–620.
18. Hammersley, J. M., & Handscomb, D.C. (1964). *Monte Carlo methods*. Methuen. London.
19. Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall.
20. Hall, P. (1992). *The bootstrap and Edgeworth expansion*. Springer-Verlag.
21. Shorack, G. R., & Wellner, J. A. (1988). *Empirical processes with applications to statistics*. New York: John Wiley & Sons. p. 639.
22. Wilks, S. S (1962). *Mathematical statistics*. New York: John Wiley & Sons. Section 11.2.
23. Kotz, S., & Johnson, N. L. (Editors-in-Chief) (1986). *Encyclopaedia of statistical sciences*. New York: Wiley & Sons, New York. Vol. 7, p. 427.
24. Angus, J. E. (1993). Asymptotic theory for bootstrapping the extremes. *Commun. Statist. Theory Meth.*, **22**, 15–30.
25. Breiman, L., Stone, C. J., & Kooperberg, C. (1990). Robust confidence bounds for extreme upper quantiles. *J. Statist. Comput. Simul.*, **37**, 127–149.
26. Caers, J., Beirlant, J., & Vynckier, P. (1998). Bootstrap confidence intervals for tail indices. *Comput. Statist. Data Anal.*, **26**, 259–277.
27. Bertail, P. (2000). Risk in food. Lecture given at INA-PG. Paris, December 2000.