Probabilistic modelling of acute dietary exposure of pesticide residues and other contaminants

Andrea Zentai

PhD thesis summary and statement

Food Safety Risk Assessment Directorate
National Food Chain Safety Office
NÉBIH ÉKI

2015

PhD School/Program

Name: PhD School of Food Sciences

Field: Food Science

Head: **Dr. József Felföldi, PhD**
Department of Physics and Control
Faculty of Food Science
Corvinus University of Budapest

Supervisor: **Dr. habil Árpád Ambrus, CSc**
honorary professor, retired chief scienfic advisor

The applicant met the requirement of the PhD regulations of the Corvinus University of Budapest and the thesis is accepted for the defence process.

..... **.....**
Head of PhD School Supervisor

1. INTRODUCTION

The application of pesticides is unavoidable in order to have food of sufficient quality and quantity for the increasing population of the world. When used according to the rules, the pesticide residues are not the biggest threat to our food safety, in spite of the fact that a major part of the population is still the most worried about pesticide residues.

Risk assessment of pesticide residues intake of the whole Hungarian population has not yet been performed, although appropriate data are available. The significance of this work is further underlined by the abovementioned consumer concern. I intend to fill this gap with my work by further developing and adapting the novel estimation procedures published in the international scientific literature.

To calculate the intake by point estimation, four different cases are applied according to the unit size and processing status of a crop, which are described by the so called IESTI (international estimated short term intake) equations. The IESTI equation, however, can not deal with cases when the residues of a chemical can be found in several consumed foods, or the frequency of consumption. Probabilistic methods can be the solution for this. This approach can supplement, refine the deterministic method, according to a tiered approach, when the result of the latter could not clearly exclude the risk.

According to conclusions of recent publications testing the probabilistic guidance of the European Food Safety Authority (EFSA), more experience and a realistic procedure would be needed, taking into account elements of the optimistic and pessimistic approaches recommended by EFSA, which can be still be considered conservative enough, but not so much as the current pessimistic approach.

I use probabilistic estimation in my work to calculate the exposure of harmful chemicals/contaminants of food (deoxynivalenol, morphine in poppy seed, captan) and cumulative exposure of several pesticide residues (organophosphorus compounds).

The MCRA is a well-known European exposure assessment software, which was being developed parallel to my work. New developments include the recommendations of the EFSA guidance, and elements of cumulative and aggregated (several intake routes) exposure assessment. The detailed operational algorithm of the models is not available, and no reference can be found in recent publications for the application of some alternative approaches (special elements to model unit mass variability and cumulative intake), which would be desirable to incorporate in a risk assessment software. The system is in fact a black box using up strictly defined input parameters, which would necessitate the conversion of our national food consumption database structure, or perform a new consumption survey according to the structural rules of MCRA.

Application of a special national software would be desirable therefore, which builds upon the national databases, and the user is familiar with the background of input data and their quality, and the operation of the program step by step. In my work I developed the algorithm of such a software, and make suggestions to its development.

2. OBJECTIVES

1. Based on the available national data and the probabilistic methods recommended by the international literature and guidance documents, the development of methods to perform exposure assessment for selected national cases and the calculation of intake in case of

- deoxynivalenol (DON) intake from white bread consumption
- acute morphine intake from poppy seed consumption.

2. Studying the factors affecting the acute exposure assessment of pesticide residues (unit mass, unit residue variability, residue values below the limit of detection, and number of bootstrap iterations), on the example of captan exposure from apple consumption.

3. Development of databases of unit masses and variability factors to perform the refined exposure assessment.

4. By taking into account the results of international research, the development of a probabilistic method based on the national databases to determine the cumulative exposure to several chemicals with the same toxicological mechanism of action. Application of the method to estimate the cumulative acute exposure of the whole population and selected age groups to organophosphorus compounds.

5. Development of the logical system plan of a complex computerized system, which is able to estimate the short-term (acute) or long-term (chronic) exposure of the whole population or selected population groups to any chemical (particularly pesticide residues) single or cumulative, from the whole diet, or selected foods consumption, based on the national individual consumption data and the laboratory measurement results.

6. The tasks performed in Excel macros under objectives 1, 2, and 4 (exposure assessment of DON, morphine, captan and organophosphorus compounds) will be applicable to check and validate the correct functioning of the complex computerized system realized based on the logical system plan which I elaborated. For this purpose I express the input parameters according to the format of the system plan.

3. MAIN DATA AND METHODS USED

3.1. The consumption data are derived from the food consumption survey performed by HFSO (Hungarian Food Safety Office) in 2009 by 3-day dietary record method. Altogether 4992 peoples 3-day consumption was recorded. In case of DON and morphine, data from the former survey of 2003 (OLEF) were used. This survey consists of 3-day consumption data of 1360 adults.

3.2. The Hungarian official pesticide residue analytical results are available in an electronic database. The user can extract the data from the database by several queries. Data can be sorted according to the time of analysis, crop, residue, or the type of survey. The database contains the code of the sample, chemical, analysing laboratory as well as the limit of detection of the method and the maximum limit of the chemical according to the regulation. Further contaminant concentration data were derived from official measurements and private laboratories.

3.3. As part of my work I also collected data for the unit mass of fruits and vegetables. With my colleagues we weighed about 50 types of fruits and vegetables on a digital scale, sometimes hundreds of pieces in several supermarket stores. Knowledge about the mass distribution of produce enables to model the unit mass and variability probabilistically in acute exposure assessment.

3.4. The applied probabilistic methods and further specific data are presented at the relevant sections.

4. RESULTS

4.1. DON intake of Hungarian adult consumers derived from white bread consumption

For this assessment, data from a special survey were applied, which was initiated in 2009 by HFSO because of the high mycotoxin contamination found in cereals. In addition, analytical results from the regular monitoring of wheat flour, and DON concentration data from pre-export control of wheat were also taken into account. This meant altogether 176 samples of wheat flour and 147 samples of wheat grain.

White bread consumption data were extracted from the 2003 OLEF survey. All consumption days were taken into account for the probabilistic procedure.

The change in mycotoxin contamination resulting from processing was modelled by using the processing factor (ratio of DON concentrations in the final and the raw product). To count for the reduction of DON concentration during wheat grain milling, the median of the processing factors available in the literature, 0.471 was applied. For bread baking, 0.7 was used as a processing factor, based on the recipe stating 700 g flour was needed for 1 kg of bread. DON concentration was calculated from the analytical data of wheat grain and flour, using the relevant processing factors. Two types of probabilistic methods were used.

During procedure „A", 200,000 samples were taken with replacement from both the bread consumption and the DON concentration data, and these were multiplied, thus resulting in 200,000 intake values. During procedure 'B', all consumption and DON concentration data were multiplied.

Table 1. Comparison of intake calculated by point estimate and probabilistic method

	Wheat flour	**Wheat grain**
Average consumer (point estimate)	0,282 µg/kgbw	0,490 µg/kgbw
High consumer (point estimate)	**0,606 µg/kgbw**	**1,053 µg/kgbw**
Median (probabilistic)	0,1-0,15 µg/kgbw	0,1-0,15 µg/kgbw
95. percentile (probabilistic)	1,0 µg/kgbw	2,1-3 µg/kgbw
Exposures above PMTDI	~5%	~15%

- As a methodological conclusion from the comparison of the results from the two methods (A and B), the resulting distributions are independent of the methods, so by sampling enough data compared to the basic population, the result will be reliable.
- The provisional maximum tolerable daily intake (PMTDI) of DON is established by JECFA as 1 µg/kgbw. The point estimate of the high consumer (95 percentile of consumptions) was about 60% of PMTDI, based on the average contamination level of wheat flour, but it was around the PMTDI, based on wheat grain contamination. Therefore, based on average conatmination of wheat, the intake of a person consuming white bread at the 95 percentile level of daily consumptions can reach up to the provisional maximum tolerable daily intake (Table 1). These higher intakes derived from wheat grain can be explained by the originally measured relatively higher concentration values, compared to wheat flour.
- The medians and 95 percentiles of the intake distributions (based on wheat grain data) ranged between 0.1-0.15 µg/kgbw and 2.1-3 µg/kgbw. In case of substitution of wheat flour data by LOQ (limit of quantification) and LOD (limit of detection) or using the original data, the 95 percentile intakes were at around the PMTDI.
- Concluding from the probabilistic calculations, the DON intake from only white bread consumption exceeded the PMTDI in case of 5-15% of Hungarian consumers in 2008-2009 (depending on the applied data and methods). The most exposures above PMTDI were based on the calculations with wheat grain contamination data.

4.2. Probabilistic estimation of acute morphine exposure derived from poppy seed consumption

566 and 171 poppy seed analytical results measured by the National Institute for Food Hygiene and Nutrition between 2001-2006 and Central Agricultural Office between 2007-2010 were used in this study. Beside the 2003 OLEF survey, the newer consumption data of 2009 were included in the study too. In the first case, 79 consumption days were recorded (1,94% of all), and 327 in the latter (2,18%). A great advantage of the 2009 consumption database is that it contains children consumption data, and of the 1010 involved children, they consumed poppy seed on 85 days. I could also conclude from the 2009 database that in 65% of the cases the consumption concerned poppy seed sprinkled on pasta, and in 35% it concerned poppy seed in cakes. Based on literature data, Pf=0.71 was used for grinding (29% reduction), and Pf=0,22 was used as the combined processing factor for grinding and baking. The reduction of morphine content due to washing of poppy seed was not taken into account as it is not common practice in Hungary to wash poppy seeds.

The morphine intake was calculated first by multiplying the 97.5 percentiles of poppy seed consumption and concentration data as point estimate, both separately and combined from the two consumption and concentration databases, to characterize potential trends. Four methods were used for the probabilistic assessment.

a) All consumption data were multiplied by all concentration data (estimation based on empirical data);

b) 200,000 or 500,000 consumption and concentration data pairs, generated from the fitted distributions, were multiplied (parametric modelling);

c) The combined 737 morphine concentrations and 406 consumption data were bootstrapped for 10,000 times (random sampling with replacement with the same number from the original data) and the exposure was calculated from these databases (estimation based on empirical data);

d) Performing procedures b) and c) when taking into account the effect of processing (concentrations reduced by the processing factor). According to the ratio 65-35 (raw and baked forms), the values were sampled with random sampling from the 2009 consumption data, i.e. the relevant processing factor was matched to the baked poppy and sprinkled poppy consumption separately.

Table 2. Comparison of results of point and probabilistic estimate

	Point estimate	Probabilistic estimate, Exposures above ARfD
without processing factor	~116,7 $\mu g(kgbw)^{-1}day^{-1}$	~20%
with processing factor	~78,64 $\mu g(kgbw)^{-1}day^{-1}$	~10%

- Generation of 200,000 or 500,000 values from the fitted lognormal distributions resulted in different patterns. In case of consumption data, the 200,000 simulation was satisfactory to characterise the highest consumption, but in case of morphine data (having more experimental data), the highest value from the 200,000 simulation was still lower than the maximum measured concentration, therefore, the 500,000 simulation was necessary. Based on these results, in the upper tail region of concentration data, generation of parametric distributions of a higher data number could be needed.
- Comparing the distributions resulting from the empiric and parametric models, up to about the 90 percentile intake the results were in the same range, however, towards the 99.9 and 99.99 percentiles exposures the results from the empiric approach was more than two times higher, even in case the highest generated values from the parametric distribution were above the highest experimental value.
- The 95% confidence interval derived from the bootstrap procedure was narrowest at the 50 percentile (median) and getting larger at the lower and higher percentiles. The calculated exposures confidence interval up to about the 99 percentile can be described by the normal distribution but the distribution is getting scattered above this percentile.

•The intake calculated by point estimate ranged between 73.1 and 116.7 µg/kgbw/day depending on time periods (processing effect not yet taken into account), i.e. the lowest estimated value is still much higher than the 10 µg/kgbw acute reference dose established by EFSA (Table 2). As children consumption data were available only from the 2009 survey, the 2007-2010 and the 2001-2010 combined morphine data populations were matched to these consumptions, resulting in 164.7 and 150.2 µg/kgbw/day intakes (processing effect not yet taken into account).
•Taking into account the effect of processing, the point estimate of adults intake was 78.64, and of children was 116.9 µg/kgbw/day.
•Concluding from the intake distribution estimated by probabilistic method based on data described by lognormal distributions (not yet considering the effect of processing), the intake of the 97.5% of poppy seed consumers was 25.6-34.7 µg/kgbw/day at most (depending on time periods), while for children this value was 59.1 µg/kgbw/day (concerning the 2007-2010 concentration data). The exposure estimated based on the empiric approach was higher, 38.8 µg/kgbw/day at the 97.5 percentile for adults.
•Considering the reducing effect of processing on morphine levels, the 97.5 and 99 percentiles were lower, between 18.3-25.4 µg/kgbw/day and 25.6-47.4 µg/kgbw/day, and 32.9 and 66.4 µg/kgbw/day for children.
•The proportion of intakes above the acute reference dose was about 20% for adults, and about 10% when considering the reducing effect of processing (based on the fitted lognormal distributions which resulted in lower results) (Table 2.).

4.3. Factors affecting the probabilistically estimated acute exposure on the example of captan intake derived from apple consumption

Regarding the residue results, the time periods of 2005-2009 and 2010-2011 were taken into account separately and combined as well. According to pesticide usage records I assumed that about 20% of apple orchards were treated with captan. To estimate the acute exposure therefore, only the measured captan data and non-detects reaching up to 20% of all results were taken into account, for which the relevant data below LOQ were substituted with the maximum likelihood estimation (MLE) method. I also studied the effect of non-detects by substituting with LOQ and 0 as well.

I developed a procedure to take into account the apple elements mass and residue variability as distributions, thus further refining the EFSA basic assessment approach towards the refined assessment approach. The short term captan intake of one apple consumption day was calculated by equation (1).

$$ESTI_{nk} = ((R_k * \upsilon_{i1} * m_1) + (R_k * \upsilon_{i2} * m_2) + \ldots (R_k * \upsilon_{il} * m_L)) / bw_n \qquad (1)$$

where n denotes the consumption day, R_k is the average residue measured in a composite apple sample of K elements, υ_i denotes a random unit variability factor (unit residue/ average residue content) for a given apple, m denotes the unit mass of the given apple.

The elements of the sum mark the apple pieces eaten by the consumer. The sequence is continued until the apples eaten equal the recorded daily apple consumption of the given consumer, i.e. until the sum of m_1, m_2, ... m_L masses equals the consumption. The last m_L piece can be only a fraction of a whole piece of apple, with which the sum of selected apple masses will be equal to the recorded daily consumption. The unit mass and unit variability of consumed apple pieces were selected by random sampling with replacement.

To study the effect of different factors, I used several procedures as follows.

a) basic exposure calculation: every consumption value was multiplied by every composite sample residue concentration.

b) effect of random selection of m and υ: The daily exposures were calculated according to equation (1) by randomly selecting from the variability and apple mass values. The procedure was repeated 100, 200, 500 and 1000 times.

c) effect of bootstrap: the basic exposure calculation (method *a*) with bootstrap repeated 100, 200, 500, 1000 and 10000 times.
d) combined effect of random selection of *m* és *υ* and bootstrap: bootstrapping the residue concentration and consumption data 100, 200, 500, 1000 and 10000 times, then for each iteration, applying procedure b) based on equation (1).

Table 3. Selected percentiles of intake distribution estimated based on 2010-11 data (10,000 bootstrap)

Percentile →	95	98	99	99.99
General population (µg/kgbw/day)	4.8	10	16	133
Women of age 15-45 (µg/kgbw/day)	4.01	8.1	12.0	53.8
Women of age 15-45 (ARfD %)	1.3	2.7	4.0	**17.9**

- Comparing the distribution patterns derived from 100, 500 and 1000 iterations, it could be concluded, that above 100, increasing the number of iterations changed the relative cumulative probability distribution of exposure only maximum 1-2 percent.
- Up to (and under) the 95 percentile the exposure calculated by taking into account the unit mass and variability was lower, however, at about the 95 percentile this tendency changed and higher exposures resulted when taking into account the unit mass and variability. Although towards higher percentiles the two data populations gradually differ, the 95% confidence intervals overlap.
- In case of the two different time periods (2010-2011 and 2005-2011), comparing the results following 1000 bootstrap it can be concluded, that at the higher percentiles (90-99.99) the results were about 26-31% lower in the latter case. An explanation for this could be the higher proportion of lower residue concentrations in the period 2005-2011.
- My results call attention to the fact that in case the proportion of non-detected residue results are high (60-80%), then even the substitution with LOD can't be considered a conservative approach. In such cases it is more practical to use the measured values only, or perhaps those from the fitted parametric distributions on the measured values only, as a conservative approach.
- At the 95, 98, 99 and 99.99 percentiles the d) procedure, which gave higher exposures at the high percentiles, resulted in 4.01, 8.1, 12.0 and 53.8 µg/kgbw/day for the women of child bearing age, while for the general population these values were higher, 4.8, 10, 16 and 133 µg/kgbw/day. Compared to the acute reference dose of 300 µg/kgbw/day established by the JMPR, the exposure of the women of child bearing age at the 95, 98, 99 and 99.99 percentiles was 1.3, 2.7, 4.0 and 17.9%, so according to these results, the application of captan on apples does not pose an acute health concern for the Hungarian consumers (Table 3).
- The point estimate exposure of the general population (95 percentile of apple consumption multiplied by the 97.5 percentile of captan concentrations) based on the 2010-2011 data was 17.4 µg/kgbw/day.

4.4. Probabilistic cumulative acute intake assessment based on the example of organophosphorus compounds exposure

Based on the sample code the residue concentrations measured in the same sample can be connected, and, based on their toxicological reference values, can be expressed as equivalents of the selected index compound. Acephate was chosen as index compound, and after the acephate equivalent concentrations were calculated for each sample (equation (2)), these values were used for the daily exposure assessment.

$$Req_{acephate} = R_{acephate}*1 + R_{azinphos\text{-}methyl}*RPF_{azinphos\text{-}methyl} + R_{chlorpyrifos}*RPF_{chlorpyrifos} + \ldots\ldots \quad (2)$$

Plant based foods were selected from the consumption database. These commodities were then connected to the commodities for which there were measurements in the residue database. In some cases several commodities from the residue database could be connected to the consumption database commodities (eg. lemon and green lemon as measured to the lemon as consumed). There

were finally 23 relevant organophosphorus compounds residue data measured in 64 relevant commodities.

Zero (<LOQ) residues were not taken into account for the basic assessment. To further study the effect of residues below the detection limit (LOQ=LOD), I calculated the intake after substituting these as well, with 0 and LOQ on a randomly selected subpopulation of 1000 consumption days, as a representative sample of the consumption database. To reduce operation time of the computer (as the calculations are performed by macros developed in Excel), not all < LOQ values were substituted, but the positive measurements were supplemented with 0 or LOQ values of twice the amount, according to the 66% ratio of non-detected residues which fits the international ratio approximately.

The cumulative exposure of one consumption day was calculated through the following steps. The consumption of the first food on a given day was multiplied by all the relevant acephate equivalent residue values. Similarly, the next food consumption is multiplied by the relevant residue values. Then, the results from the first and second foods are added in all combinations. Going on to the next food, the consumption values are multiplied by all relevant acephate equivalent concentrations, and added in all possible combinations to the sums derived from the preceding steps.

The percentiles of the summed exposures for each consumption day were calculated. For the acute exposure estimation, I studied the high (97.5, 98, 99, 99.9 and 99.99) percentiles and medians at each day. By characterizing these relevant percentiles throughout all the consumption days, the cumulative exposure of Hungarian consumers could be characterized.

Table 4. Cumulative exposure (µg/kgbw) of age groups (based on the measured data)

% of consumers	0-1 yrs	1-2 yrs	3-9 yrs	10-17 yrs	18-64 yrs	65-101 yrs
99	46.09	81.10	48.09	28.34	18.95	14.19
99.9	53.80	95.27	146.14	57.07	41.15	22.90
99.99	54.57	97.41	183.93	61.08	50.52	29.68
No.	51	239	916	1354	8858	1929

No: Number of calculated daily exposures

- The higher percentiles of the 99 percentile distribution could not be calculated with NIST method in all cases depending on the number of consumers. To characterize the exposure distribution the calculations were performed with Harrel-Davis and Excel, too. There was no significant difference between the results, so further on, I used the Excel to calculate the relevant percentiles of the consumer exposure (Table 4).
- Towards the lower percentiles the substitution methods resulted in increasingly lower exposure values, than if only the measured values were taken into account. The relative reduction in exposure estimates are presented in Table 5.

Table 5. Effect of measurements below the detection limit on the estimated exposure

	0 substitution		LOQ substitution	
Whole population (%)	**P0.99**	**P0.999**	**P0.99**	**P0.999**
P0.9999	97.7%	99.4%	97.7%	99.4%
P0.999	89.0%	96.8%	88.9%	96.9%
P0.99	94.7%	83.1%	94.7%	82.8%
P0.98	86.4%	97.9%	86.3%	97.6%
P0.975	87.3%	98.2%	87.3%	98.2%
P0.5	55.0%	68.2%	55.7%	68.7%

- According to my results, the exposure estimation based on the measured values only can be considered conservative, compared to the cases when the values below the detection limit were substituted with either 0 or LOQ. The calculated exposures therefore would be lower if these

residue measurements were taken into account. As the higher percentiles of exposure are relevant for the acute assessment, the calculation based on only the positive results is acceptable.

- The 99.99 percentile of daily exposure was below the acute reference dose of acephate (100 μg/kgbw/day) for 99.95% of the consumers. Concluding from the 99.99 percentile daily exposures, $4.7*10^{-8}$ (0.047 of a million) proportion of the population would have exposures above the ARfD, maximally 2.15 ARfD (99.953%<ARfD). The occurrence of exposures at the level of ARfD has low probability; based on these data the probabilities of occurrence of the five highest exposures are between 10^{-7}% - $2*10^{-4}$%. Considering the average consumer by looking at the median of the 98 percentile daily exposures, the intake is lower than 1% of the ARfD.
- The exposure of the age group 1-9 years is higher than the older age groups, particularly for the 3-9 years age group; from consumers only, 40 children of a million could have intakes above the ARfD. As the ratio of detected residues in the consumed produce/foods ranged between 0.1-1.6%, the children of 3-9 years presumably consumed foods containing acetylcholinesterase inhibitor pesticide residues in such a ratio. Assuming the 1.6% treatment ratio as the worst case, the intakes above ARfD would occur only in case of a fraction of children (0.64 children of a million).
- I generally concluded, that the cumulative exposure to organophosphorus residues from plant based food consumption did not pose a health concern for the general population, including the age group of 3-9-year-old children.

4.5. Logical system plan for the probabilistic modelling of the consumer exposure based on the national data structure

To provide a complex solution for our national dietary risk assessment tasks, I developed the plan of a computerized program operating with the principles of probabilistic modelling, which basically builds upon the pesticide residue intake assessment methods, but can be extrapolated and applied to other contaminants, too. In the following sections I present the basic steps specifically for the pesticide residues.

4.5.1. Intake of one day from one food component and one concentration value

The two main input data needed for the exposure assessment, the measured value ($R_{ka_{1j}'}$) (concentration) of the *'k'* parameter (contaminant) in *'a_1'* food component and the consumption (F_{ia_1}) in terms of *'k'* parameter need to be expressed. The calculation is then carried out according to equation (3) (EDI = estimated daily intake).

$$\mathrm{EDI}_{ia_1k_j} = \frac{R_{ka_{1j}'} \times F_{ia_1}}{\mathrm{kgbw_i}} \qquad (3)$$

The value of $R_{ka_{1j}'}$ is chosen by the program from the data population of measured values, or from the modified version where the values below the detection limit (< LOD) have been substituted, *j* denotes the results of measured samples of the food component, and kgbw is the body weight of the consumer.

4.5.2. Expressing the consumption of the food component

The quantities of the consumed foods are expressed in their given component, for which the measurement of the residue concentration was performed. For instance, if the consumer exposure is calculated based on the malathion residue of flour, then the flour content of all the foods consumed on that day without any measurements performed explicitly for malathion (e.g. bread, bread rolls, soup pasta, etc) has to be added. The consumed total quantity (F_{a_1}) of the first relevant food component (a_1) is calculated by equation 4.

$$F_{a_1} = \left(f_{a_1} \times K_{a_1 k} \times 1\right) + \left(f_{t_1} \times H_{a_1 t_1} \times Pf_{a_1 t_1 k} \times K_{a_1 k}\right) + \left(f_{t_2} \times H_{a_1 t_2} \times Pf_{a_1 t} + \ldots \right. \quad (4)$$
$$= F'_{a_1} + F'_{t_1} + F'_{t_2} + \ldots$$

f_{a_1} is the consumed amount of the edible portion of the raw commodity (e.g. orange pulp)

$K_{a_1 k}$ is the ratio of the concentration of the *'k'* residue in the edible part of the commodity and the whole commodity (of which the measurement was performed). For example the kernel of peach or the peel of melon is not consumed, however, the residue is determined for the whole commodity (*a_1'*) according to the relevant Codex standard and the corresponding EU regulation. $K_{a_1 k}$ factor is dependent of the food component and the measured parameter. The residue values are the average concentrations of composite samples.

f_{t_1}, f_{t_2}, etc. are the consumed amount of the processed foods containing the *'a_1'* component (e.g. orange juice, orange jam)

$H_{a_1 t_1}$ is the ratio of the *'a_1'* component in *'t_1'* food. E.g. the amount of apple in a 40% apple juice is 0.4.

$Pf_{a_1 t_1 k}$ denotes the change in the residue concentration during the processing of the component *'a_1'* into *'t_1'* food product. The Pf_{atk} depends on the raw commodity, the processed food, the way of processing and measurand. The K_{ak} and Pf_{atk} factors are only relevant together in the rare cases where the edible portion is further processed.

In case when consumption and concentration data are only available for the *'a_1'* component, then the equation (4) is simplified to the first element.

The consumption database contains consumptions of components (*a*) which can not be further degraded, and composite foods (*t*). The elements of the sum (F'_{a_1}, F'_{t_1}, F'_{t_2}, etc. in equation 4) are the subintakes from the component and from the composite/processed foods.

4.5.3. Intake from all the food components consumed on a day

To calculate the total daily intake the program performs the steps described above for all *'a'* food components (e.g. apples, oranges, potatoes, etc.) from the given consumption day, i.e. similarly to white flour, the composite/processed foods made of e.g. apples, oranges, potatoes, etc. need to be expressed accordingly. The calculation is made according to equation (5).

$$EDI_{iak} = \frac{1}{kgbw_i} \sum_{a=1}^{A} [F_{ia} \times R_{ka'i}] \quad (5)$$

A is the number of different consumed food components on day *'i'*;

$R_{ka'i}$ denotes the concentration of *'$k_{a'}$'* parameter (e.g. residue) measured in *'a''* commoditites corresponding to the food components (*$a_1 \rightarrow A$)* consumed on day *'i'*.

4.5.4. Intake estimation taking into account all relevant concentration data

The program repeats the calculation for all the *'j'* values of the *'k'* residue measured in the commodities corresponding to the food components consumed (equation (6)). Accordingly, the resulting exposure estimates equal to the highest number of residue measurements ($S_{a'}$) from the commodities corresponding to food components consumed on the given *'i'* day.

$$i(EDI_{ki}) = \frac{1}{kgbw_i} \sum_{a=1}^{A} \left[F_{ia} \times \prod_{j=1}^{S_{a'}} R_{ka'j} \right] \quad (6)$$

i(EDI$_{ki}$) marks the calculated exposures for day *'i'*.

4.5.5. Taking into account unit residue variability for acute assessment

In cases during acute exposure assessment, when the f_{a_1} consumption of equation (4) exceeds the lower value of the unit mass distribution of the commodity, then the consumer will consume more than one unit of the concerned commodity. In such cases the different residue concentrations of the units need to be considered for exposure assessment. The *'$f_a^{\dagger}$'* consumption which is modified by unit residue variability is calculated by equation (7).

$$f_a^{\dagger} = \left(\upsilon_{a_1'} \times m_{a_1'}\right) + \left(\upsilon_{a_2'} \times m_{a_2'}\right) + \cdots + \left(\upsilon_{a_L'} \times m_{a_L'}\right) \tag{7}$$

$\upsilon_{a_1'}$ is the unit variability factor dependent on food component *'a_1'*, $\upsilon = \frac{R_{unit}}{\bar{R}}$

Where R_{unit} marks the residue content of unit elements in a lot, and $\bar{R}$ is the average residue content of the lot;

$m_{a_1'}$ is the unit mass of food component *'a_1'*, $m_{a_L'}$ is the last selected unit or its corresponding proportion. Both values are selected by the program randomly with replacement in a way such that the condition f_a=$\Sigma m_{a'}$ is fulfilled. For the further calculations, the $f_a^{\dagger}$ value is substituted into equation (4) (the term representing the consumption of the raw commodity). The further terms of the equation concern the processed food products, where the residue variability is not relevant.

4.5.6. Calculation of cumulative exposure

To calculate the cumulative exposure in a given commodity the summed concentrations of the chemicals with the same mechanism of action need to be considered together. In such cases, the concentrations of different residues expressed in equivalents of the selected reference chemical ($R'_{k_1a'_j}$, $R'_{k_2a'_j}$, $R'_{k_3a'_j}$, *etc.*) need to be put in place of the $R_{ka'_j}$ value (equation 8).

$$R'_{k_1a'_j} = R_{k_1a'_j} \times RPF_{k_1} \tag{8}$$

$R'_{ka'_j}$ is the equivalent concentration expressed in terms of the reference chemical in the given sample.

The relative potency factor (RPF) is based on literature data, or can be calculated as the ratio of the acute reference doses (equation (9)).

$$\mathrm{RPF_k} = \frac{\mathrm{ARfD_{ref}}}{\mathrm{ARfD_k}} \tag{9}$$

The $R'_{ka'_j}$ values are multiplied by the sub-consumptions (F'_{a_1}, F'_{t_1}, F'_{t_2}, etc.) calculated in equation (4) to estimate the cumulative exposure derived from *'a_1'* food component (equation (10)).

$$EDI_{ia_{1j}} = R'_{k_1a_1'_j} \times \left(F'_{a_1k_1} + F'_{t_1k_1} + F'_{t_2k_1} \ldots\right) + R'_{k_2a_1'_j} \times \left(F'_{a_1k_2} + F'_{t_1k_2} + F'_{t_2k_2} \ldots\right) + R'_{k_3a_1'_j} \times \left(F'_{a_1k_3} + F'_{t_1k_3} + F'_{t_2k_3} \ldots\right) + \ldots \tag{10}$$

The calculation is again performed for all *'j'* measured values (from the further j_1, j_2, etc. samples) corresponding to the food commodity *'a_1'*. Similarly to the procedure applied when there was only one measured parameter, the number of resulting values in a row equals the number of analysed *j* samples from the commodities corresponding to food component *'a_1'*.

The calculation is then repeated for all the further *'a_1'*, *'a_2'*, etc. components consumed on the day. The number of results corresponding to the *'a_1'*, *'a_2'*, etc. food components is different depending on the measured sample number *'j'*.

4.5.7. Summing the exposure for the given consumption day

The program calculates the summed exposure of the given *'i'* consumption day by adding the intakes calculated for each *'a'* food component in every combination, as already presented in section 4.4.

4.6. Calculation of exposure for individual cases

The examples of DON, morphine, captan and cumulative organophosphorus intake assessments can be utilized to test the functioning of the program. For this it is necessary to fill the input files of the software with the relevant data and select the parameter conditions, similarly as was applied in the given cases.

4.7.Conclusions of the national risk assessment studies

4.7.1. Comparison of applicability of the used probabilistic methods

I basically used three types of probabilistic methods during the case studies. According to literature, the distributions of residues/contaminants in most cases approximate the lognormal distribution, which was supported by my findings as well. In case of morphine exposure estimation, the empiric procedure resulted in more than twice the values than the results simulated from the fitted lognormal distributions at the 99.9 and 99.99 high percentiles. The assumed reason for this was that the data simulated from the fitted lognormal distribution was not enough to cover the highest values of the basic data, or the distribution could not exactly describe the distributions of concentration and consumption data. The lower frequency of occurrence of extreme high values is a reason to study their realism and the number of data simulated from the fitted distribution, whether they fit well the original data at the higher percentiles (not much higher, but do not underestimate).
In case of the empiric procedures I multiplied all data with all data, and also used the procedure of random sampling. In the latter case the minimum required sample number can depend on the given case, therefore the EFSA recommends to continue the simulation until the distribution curve is not changing. In case of morphine exposure estimation, from the 200,000 and 500,000 simulation the latter provided more appropriate results.

Multiplying all the data by all the data shows all the possible intake values based on the available data. The procedure is reliable, when sufficient data are available. The advantage of parametric modelling comes forward in case of having only a few data. Naturally, in case of a few data, determining the relevant parameters of a parametric distribution is uncertain, which appears in the uncertainty of the estimated exposure.

4.7.2. Applicability of bootstrap

The bootstrap procedure can not give more exact results than the procedure of multiplying all the data, but makes it possible to characterize the uncertainty caused by random sampling from the original data, by providing the confidence intervals around the cumulative relative frequency curves. This interval is narrowest at the medians and towards the lower and higher bounds it is getting wider, meaning the possible occurrence of data can be expected in a wider interval.

The number of necessary bootstrap iterations is dependent on the number of original data. The iterations need to be repeated until the bounds of the resulting distribution do not change significantly. According to my results, 500-1000 iteration was sufficient. In case of captan acute exposure assessment I found that after 100 iterations the confidence intervals of the result did not change significantly.

4.7.3. Refinement with variabiliy factor and unit mass

The procedure developed for refinement with the variability factor and unit mass in case of medium sized commodities resulted in higher values above the 95 percentile, than the procedure of multiplying all the data. The reason for this can be that multiplying by the variability factor (value

often above 1) will shift the residue values and consequently the calculated exposure value higher. I recommend this procedure as a conservative approach.

4.7.4. Treatment of samples with results below the detection limit

In case of residue databases the ratio of non-detects can be 80-90%. I concluded that the estimated exposure is lower when the values below the limit of detection are considered by substituting with 0 or LOQ. In case of the two substitution scenario (0 and LOQ) there were no significant differences between the calculated cumulative exposures at the high percentiles.

Considering the measurements in the treated ratio of the commodity could further refine the estimate. It is also very important to study the realism of the applied residue databases, and considering only those values which are justified (e.g. authorized, or efficient in the given commodity). According to these conclusions, counting with the measured values only is considered more conservative than the substitution with LOQ.

4.7.5. Exposures exceeding the toxicological reference value

To inform decision makers the EFSA guidance recommends to communicate the ratio of exposures exceeding the reference value (ARfD, ADI) of a million cases. These data are easier to interpret than the occurrence probability of an exposure above a given percentile.

The DON exposure calculated from wheat grain data exceeds the PMTDI of DON in ~112,700 cases of a million, while that calculated from wheat flour exceeds the PMTDI in ~33,800 cases.

Estimated from the lognormal distribution fitted to morphine concentrations and taking into account the effect of processing, the intakes exceed the acute reference dose of morphine in 100,000 cases of a million. Taking into account the proportion of poppy seed consumptions in the consumption database, the ratio derived for the whole population is 2180 of a million.

Based on the intake assessment of captan, the calculated exposure exceeds the acute reference dose in less than 100 cases of a million. Further refining with the ratio of apple consumptions and apple treatment ratio, the concerned proportion of a million is negligible (6 persons of a million).

In case of the organophosphorus compounds cumulative intake, 0.047 exposures of a million are over the ARfD, maximally $\leq$ 2.15 ARfD. In case of the 3-9 years age group, about 40 children of a million would be exposed to intakes higher than the ARfD. Refining the estimate with the 1.6% maximum ratio of detected residue values in the samples of consumed foods the risk can be considered negligible (0.64 children of a million).

4.7.6.Comparison of point and probabilistic estimate

Probabilistic estimation can be considered a refinement of the simple point estimate. The principles of the two approaches differ, while at the point estimate the result is a point value, the significance of probabilistic estimate is giving the distribution of intakes and not to highlight a given point (e.g. 97.5 percentile).

For the point estimate of acute exposure usually the 97.5 percentile residue concentration is multiplied with a high consumption (97.5 percentile). The occurrence of the recorded highest food consumption from the most contaminated food is very unlikely. Table 6 illustrates how much could be the difference between the results calculated from the 97.5 percentiles or the maximums, on the examples of DON and morphine intake assessment.

Table 6. Comparison of point estimates calculated from high percentile values and maximums

	High intake (97.5 percentile consumption * 97.5 percentile concentration)	**Maximum intake**	**Probability of occurrence of maximum intake**
DON (from wheat flour)	3.1 µg/kgbw	15.7 µg/kgbw	$4.2*10^{-4}$%
Morphine (combined period)	$1*10^{2}$ µg/kgbw	$1.5*10^{3}$ µg/kgbw	$3.3*10^{-4}$%

4.7.7.The uncertainty of assessment

I characterised the uncertainty of the exposure estimation at the given cases, by describing the confidence interval of the results and/or modelling with alternative assumptions. By considering further uncertain elements of the assessment based on the EFSA table recommended for unquantified uncertainties, I concluded that the resulting morphine exposure is an underestimation of the high exposure, because the consumption survey did not cover the main poppy consumption time periods (Easter and Christmas). In other cases there were sufficient consumption and measured residue or contaminant data. I performed the assessment with only the measured values and with those under the LOQ by different substitutions, concluding that the exposures calculated from only the measured values can be considered conservative.

5. THESIS STATEMENTS

1. I elaborated the logical system plan of a model which is able to estimate the exposure, taking into account all parameters influencing the exposure (residue concentration of edible portion of the commodity, processing factor, variability factor and unit masses).
The procedure, which enables a more complex estimate compared to some elements of the latest version of MCRA

- is aligned to the structure and data content of Hungarian databases;
- enables to analyze the sources of exposure; the contribution of different foods and the analyzed chemicals individually or cumulative to the total exposure;
- appropriate to calculate the consumer exposure not only to pesticide residues, but any chemical contaminant.

2. I established the relations of consumer exposure calculations and described with exact equations. The method can be considered a refinement of those found in literature and enables a more detailed estimation of exposure.
The basic relations of the calculations are valid for the estimation of acute and chronic exposure for different combinations of measured chemicals and food consumptions:

- one measurand – one food
- one measurand – several foods consumed on the same day
- several measurands (cumulative effect) – one or several foods consumed

3. I elaborated a model for the refinement of acute exposure, where the variability factors and unit masses are sampled randomly from the relevant distributions, according to the equation:

$$ESTI_{nk}=1/kgbw_n((R_k*\upsilon_{i1}*m_1)+(R_k*\upsilon_{i2}*m_2)+\ldots(R_k*\upsilon_{il}*m_L))$$

R_k denotes the average residue content of the composite sample, υ_i is the individual variability value of the commodity unit, m is the unit mass of the commodity. The results of the model are higher at the upper percentiles of the distribution (above the 95 percentiles), than those resulting without the application of υ_i and m. I suggest to apply this procedure as a conservative approach.

4. With probabilistic and deterministic methods I determined the Hungarian populations' exposure to deoxynivalenol (DON) derived from white bread consumption.
The results showed that the DON intake of 5-15% of Hungarian adult population exceeded the provisional maximum tolerable daily intake (PMTDI), which is 1 μg/kgbw.

5. With probabilistic and deterministic methods I determined the Hungarian consumers' exposure to morphine from poppy seed consumption.
The calculated intakes, interpreted for the whole population, and taking into account the reducing effect of processing and the ratio of poppy seed consumptions, exceeded the acute reference dose of 10 μg/kgbw established by EFSA in about 2180 cases of a million.

6. With probabilistic and deterministic methods I determined the Hungarian populations' exposure to captan derived from apple consumption.
By applying the procedure resulting in the highest exposure from apple consumption, the intake of women of child bearing age (for which age class the acute reference dose had been established), was 1.3, 2.7, 4.0 and 17.9% of the 0.3 mg/kgbw/day acute reference dose at the 95, 98, 99 and 99.99 percentiles.

7. With probabilistic method I determined the cumulative intake of Hungarian population to organophosphorus residues from plant based foods.

I concluded, that the cumulative exposure to organophosphorus residues from plant based food consumption did not pose a health concern for the general population, including children.

6. PUBLICATIONS

Publications in journals with impact factor

Ambrus, Á., Szeitzné-Szabó, M., Zentai, A., Sali, J., Szabó, I.J. (2011): Exposure of consumers to deoxynivalenol from consumption of white bread in Hungary. *Food Additives and Contaminants*, 28(2), 209-217. DOI: 10.1080/19440049.2010.540720. IF (2011) = 1.765

Zentai, A., Sali, J., Szeitzné-Szabó, M., Szabó, I.J., Ambrus, Á. (2012): Exposure of consumers to morphine from poppy seeds in Hungary. *Food Additives & Contaminants: Part A*, 29(3), 403-414. DOI: 10.1080/19440049.2011.636762. IF (2012) = 2.220

Zentai, A., Sali, J., Szabó, I.J., Szeitzné-Szabó, M., Ambrus, Á., Vásárhelyi A. (2013a): Factors affecting the estimated probabilistic acute dietary exposure to captan from apple consumption. *Food Additives & Contaminants: Part A*, 30(5), 833-842. DOI: 10.1080/19440049.2013.794977. IF (2013) = 2.341

Ambrus, Á., Zentai, A., Sali, J., Ficzere, I. (2011): Hidden contributors to uncertainty and accuracy of results of residue analysis. *Accreditation and Quality Assurance*, 16, 3-11. DOI: 10.1007/s00769-010-0721-6. IF (2011) = 1.036

Horváth, Zs., Sali, J., Zentai, A., Dorogházi, E., Farkas, Zs., Kerekes, K., Ambrus, Á. (2014): Limitations in the determination of maximum residue limits and highest residues of pesticides: Part I. *Journal of Environmental Science and Health, Part B: Pesticides, Food Contaminants, and Agricultural Wastes*. 49(3), 143-152. DOI: 10.1080/03601234.2014.857960. IF (2014) = 1.202

Publications in journals without impact factor

Zentai, A., Kerekes, K., Szabó, I., Ambrus, Á. (2015a): A fogyasztók növényvédőszer-maradékokból származó expozíciójának finomítása. 1. rész *Élelmiszervizsgálati Közlemények (Journal of Food Safety Investigations)* LXI, 3.

Zentai, A., Kerekes, K., Szabó, I., Ambrus, Á. (2015b): A fogyasztók növényvédőszer-maradékokból származó expozíciójának finomítása. 2. rész *Élelmiszervizsgálati Közlemények (Journal of Food Safety Investigations)* LXI, 4.

Conference article

Zentai, A., Sali, J., Ambrus, Á. (2013b): Almafogyasztásból származó kaptán expozíció becslését befolyásoló tényezők. Bódi Éva, Fekete István, Kovács Béla (editors): *Fiatal Kutatók az Egészséges Élelmiszerért*. ISBN 978-963-473-601-1. p. 90-95.

International conference

Ambrus, Á., Horváth, Zs., Sali, J., Zentai, A. (2012): Principles for Planning Monitoring Pesticide Residues in Agricultural Commodities. O-14 plenary lecture. *49th Annual Florida Pesticide Residue Workshop/North American Chemical Residue Workshop*. 15-18 July 2012, Florida

Lectures

Zentai, A., Sali, J., Ambrus, Á. (2013): Almafogyasztásból származó kaptán expozíció becslését befolyásoló tényezők. *Fiatal Kutatók az Egészséges Élelmiszerért, scientific conference.* 19 February 2013, Debrecen

Zentai, A., Szeitzné Szabó, M., Ambrus, Á., Szabó, I., Szerleticsné Túri, M., Sali, J. (2014): Mákfogyasztás élelmiszerbiztonsági szemmel. *Aktualitások a táplálkozástudományi kutatásokban, workshop*. 16 January 2014, Budapest