HUNGARIAN UNIVERSITY OF
AGRICULTURE AND LIFE SCIENCES

Deterministic and probabilistic modeling of aflatoxin M1 exposure of Hungarian consumers based on contamination of milk and dairy products

Theses of doctoral (Ph.D.) dissertation

Kata Kerekes

Gödöllő

2021

Doctoral School

Name: Doctoral School of Food Sciences

Discipline: Food Science

Head of School: **Prof. Dr. Livia Simon Sarkadi, D.Sc.**
professor
Hungarian University of Agriculture and Life Sciences,
Institute of Food Science and Technology
Department of Nutrition

Supervisor: **Árpád Ambrus, habil, C.Sc.**
honorary university professor
r. chief scientific adviser

....

Head of School approval Approval of Supervisor

1 INTRODUCTION

Mycotoxins are secondary metabolites produced by filamentous fungi (molds) that enter the food chain by contaminating agricultural crops intended for feed and food materials. The spread of various mycotoxins, the economic damage caused by them, and the human and animal health risks of their intake have been of concern to those working in the field of food safety for decades, both at European and international level. The risk assessments of the European Food Safety Authority (EFSA) and the German Federal Institute for Risk Assessment (BfR), in line with a number of other scientific publications, have highlighted that a portion of the population could be exposed to mycotoxins owing to the consumption of certain foods in excess of tolerable intakes.

Particular attention was paid to the group of aflatoxins, for which no tolerable daily intake could be established due to their genotoxic and carcinogenic nature. In 2007, EFSA's Scientific Panel on Contaminants in the Food Chain (CONTAM Panel) recommended that intakes of aflatoxins from different food sources should be kept to a minimum. The Commission's guidelines recommended that Member States carry out further studies on the subject and monitor the aflatoxin content of foods as a matter of priority. The proposal was followed by several research and studies in the Member States. In Hungary, a food safety assessment of mycotoxin contamination of cereal-based products has been completed and several exposure estimates for mycotoxins have been made in the recent years.

Other European countries have also taken an active role in aflatoxin research. The results of researchers at the University of Piacenza in Italy predicted that the incidence of aflatoxin-contaminated maize will increase in Europe due to climate change. In this regard, both the critically high-risk but less cereal-producing regions of southern Europe and the more northerly medium-risk but high-volume cereal-producing regions were highlighted, including the four main maize-producing countries: Romania, France, north-eastern Italy and Hungary, which in 2009 produced 73% of EU maize production in total.

As part of an international project lasting several years, based on data from the Italian dairy industry, my colleagues and I developed a sampling plan for the detection of aflatoxin M1 in raw milk that optimizes sample numbers yet effectively predict changes in contamination. The practical applicability of the sampling plan was verified through a case study demonstrating the use of the early warning system, which was developed based on the sampling plan. Furthermore, we determined the aflatoxin exposure of Italian consumers using aflatoxin M1 concentration data of more than 25,000 milk samples from

2013–2016. The results of the study drew our attention to the fact that the level of aflatoxin M1 exposure of some consumers poses a health concern. As it has been proven in recent years that if the weather conditions support it, the contamination of grain with aflatoxin can reach very high values in Hungary as well, we considered it reasonable to estimate the aflatoxin M1 intake of the Hungarian population.

EFSA has repeatedly assessed the aflatoxin exposure of European consumers and found the results to be worrying for both aflatoxin B1 and aflatoxin M1, especially in the younger age groups. Due to the extremely intensive international food trade, it can be assumed that the Hungarian population is exposed to a similar level of risk as the European population. Since the European estimates do not provide detailed information on the exposure of Hungarian consumers, it is important to have appropriate methods for accurate exposure estimation and to apply them to available domestic data.

1.1 Objectives

My aims were to estimate the average daily intake of aflatoxin M1 by Hungarian consumers using deterministic and probabilistic models and to examine the reality of the obtained results in the light of European data.

To this end, I set the following analytical and practical tasks to be solved:

- analysis of Hungarian monitoring results related to the occurrence of aflatoxins and concentration data measured in Hungarian dairy farms;
- utilizing the above data, using them directly and taking into account the distributions generated by parametric function fitting, to determine the distributions best characterizing aflatoxin contaminations;
- the establishment of a database of processing factors of dairy products included in the Hungarian consumption surveys, characterizing the change in the aflatoxin concentration during the production;
- characterization of the consumption pattern of milk and dairy products of the Hungarian population, examining the changes in consumption habits in 10 years;
- quantification of daily average consumption of milk and milk products by age group, expressed in milk equivalent, to characterize long-term exposure;
- development of deterministic and probabilistic models suitable for estimating the long-term aflatoxin intake of Hungarian consumers in the KNIME framework;

- comparison and analysis of exposure results obtained by running different models;
- characterization of the risk of aflatoxin intake in Hungarian consumer populations, based on the estimated exposures using the methods applied in international practice.

2 MATERIALS AND METHODS

2.1 Food consumption data

The calculations were performed with the data of two national, representative food consumption surveys conducted 10 years apart, so that we can also get an idea of the changes in the consumption habits.

The 2009 food consumption data originates from the national, representative, three-season food consumption survey of MÉBIH (Hungarian Food Safety Authority), which was prepared in cooperation with the CSO (Central Statistical Office). The survey, which was conducted with the participation of 4,992 persons, recorded the age, height and weight of the participants. By completing a consumption frequency questionnaire, participants provided amounts of infrequently consumed and special foods (e.g., dietary supplements). Eating habits were assessed with a dietary diary recorded on two weekdays and one weekend, which was completed with guidelines and a picture book to support dose assessment. The three-day survey provides food consumption data for a total of 14,976 consumption days, processed by dieticians and broken down into raw materials, to characterize food consumption habits.

Out of the 14,976 consumption days of the 2009 survey, a total of 11,267 milk consumption days (75.2%) were recorded, the frequency of sour cream and cream consumption was 52.8%, cheese consumption was recorded on 46.3% of the survey days, kefir or yoghurt consumption was recorded on 19.1% of consumption days.

The latest national food consumption data are from the 2018-2020 survey of NEBIH. The survey is part of EFSA's Europe-wide EU MENU, or "What's on the table in Europe?" project and was conducted in accordance with the recommended, uniform methodology. The participating persons were selected from the households included in the CSO Household Budget and Living Conditions survey.

During the program, two consumption days of 2,657 individuals between the ages of 1 and 74 were recorded with the help of dietitians. Participants reported on food consumed the previous day in person or in the form of a telephone interview. A picture book helped to judge the amount of food consumed. The survey was supplemented by a questionnaire on body weight and height measurement, as well as food frequency and physical activity, covering a normal week in the 12 months prior to the interview. The recording of consumption habits for ages 1-9 was supported by an eating diary.

Out of the 5,314 consumption days of the 2018-2020 survey, a total of 5,145 milk consumption days (96.8%) were recorded, the frequency of sour cream and cream consumption was 54%, cheese consumption was recorded on 60.6% of the survey days, kefir or yoghurt consumption was recorded on 24% of consumption days.

I classified the food categories of both food consumption and contamination data according to the FoodEx food classification system developed by EFSA. The FoodEx classification system was created to link the data required for exposure estimation.

2.1.1 Consumer age groups

In the case of both food consumption surveys, the consumption data of the Hungarian population were classified into 5 age categories (toddlers, children, adolescents, adults and the elderly), following the EU MENU (EFSA) methodology. As the Hungarian EU MENU survey did not cover the age group of infants (0-1 years), and due to the low number of subjects (26 people), this age group was not taken into account from the 2009 survey. There is a significant difference in the number of subjects in the age groups of the two food consumption surveys. In the 2009 survey, the age group of toddlers contains data of only 90 consumers, while in the age group of adults a large number of consumers can be observed compared to the other age groups. In the case of the 2018-2020 survey, the number of consumers is evenly distributed among the age groups. Almost all of the subjects in the survey (97-98%) consumed milk or some dairy product on the surveyed days of the consumption survey.

2.1.2 Aflatoxin concentration data

The AFM1 data are partly (1,288) from the 2011-2020 national monitoring survey of NÉBIH. 40% of the samples contained measurable amounts of AFM1. Most of the measurements were performed by ELISA and HPLC

methods on samples taken from milk from dairy farms, private producers and a small proportion of mixed milk available in shops. Analysis of mycotoxin data was preceded by data cleaning steps. From the whole dataset, I excluded the studies influencing the objective estimation e.g. the results of internal audits and proficiency tests. In addition to the large number of samples below LOQ (60%), there were also items with a very high level of contamination compared to the average. Values above 100 ng/kg were 110, 122, 141, 149, 150, 190, 238, 240, 252, 260, 292, 376, 513, 740 and 860 ng/kg, respectively. I was not able to check the validity of the results, but I did not see any justification for omitting them, so I used the full data set for my further calculations. I grouped the data according to the relevant food categories, I filtered out the irrelevant substance types e.g., feeds and I performed the classification of samples into FoodEx categories. I consulted with the laboratory staff regarding the dubious measurement results.

The other part of the AFM1 measurements were obtained from the studies of the joint project of the University of Debrecen and NÉBIH carried out until January 2021. Out of a total of 1,177 AFM1 results measured from milk, the number of samples above the LOQ was 672 (57.1%). These samples from 9 dairy farms participating in the project in 2019, 2020 and 2021 were analyzed by ELISA, and then samples with a concentration above 20 ng/kg were subjected to confirmatory HPLC analysis in the NÉBIH laboratory. For these samples, the results of the HPLC analysis were used for the calculations.

Comparing the relative frequency distributions of the NÉBIH and DE analytical results, I found that, with the exception of one outlier in the NÉBIH data (19 ng/kg), the frequency of AFM1 concentrations in the LOQ-70 ng/kg range was very similar in the two datasets, and this justifies the joint evaluation of the data. The relative frequency of samples containing AFM1 above 70 ng/kg in the NÉBIH studies was <0.5%.

A limiting factor in the risk assessment of aflatoxins was the lack of data. Following a recommendation from EFSA, food categories for which the number of positive samples does not exceed 25 or for which the proportion of samples below the limit of determination is greater than 80% should be excluded from the risk assessment. Regarding the AFM1 results, only milk tests met this criterion, while the number of tests for processed dairy products was very small.

Therefore, for processed dairy products, I could not take actual analytical results into account for exposure estimation. Instead, I calculated values derived from AFM1 concentration data measured in milk, taking into account

the processing factors of dairy products, for which I used the minimum, median and maximum values of the literature data for each food category.

2.2 Deterministic method

During the risk assessment, using the recommended tiered approach, I first determined the average exposure of the Hungarian population using a deterministic (semi-parametric) method. For this, I used the average aflatoxin M1 concentration data measured in milk. The values of the concentration data below the LOQ were taken into account with the value of the imputed (generated by distribution) data equal to the number of analytical results. The imputation was performed using parameters describing the lognormal distribution fitted to the values above the LOQ, taking into account the ratio of the values below the LOQ. For food consumption data, I used the OIM (Observed Individual Means) method recommended for long-term estimation.

First, I converted all milk and dairy product consumption data to milk equivalent using the processing factors specific to the given food category (Equations 1 and 2).

intake of $e_1, \ldots, e_j$ foods expressed in g/kg bw *(B)* expressed in milk equivalent on a given *(n)* cunsumption day:

$$B_n = \frac{\sum_{e=1}^{j}(m_e \times F_e)}{bw_n} \qquad \textit{(Equation 1),}$$

where

m_e = mass (g) of the consumed *e* food on n_i consumption day,

F is the processing factor of *e* food,

bw is the body weight of the person belonging to the given day of consumption,

and

$$F_e = \frac{C_{AFM1_e}}{C_{AFM1_{milk}}} \qquad \textit{(Equation 2)}$$

where $C_{AFM1_{milk}}$ is the concentration of AFM1 in the milk used to prepare the e food, C_{AFM1_e} is the value calculated from the minimum, median or maximum results in processed foods obtained in different experiments.

The obtained total intake values per consumption day were expressed in kg/kg bw. Multiplying the consumption amounts by the average AFM1 concentration (ng/kg), I calculated the exposure values for each consumption day (ng/kg kg bw/day). I averaged the intake values of the consumers belonging to the consumption days - 2 in the case of the 2018-2020 survey and 3 in the case of the 2009 survey. I summarized the results by consumer age groups for both food consumption surveys.

Based on the obtained exposure values, I used the margin of exposure (MoE) approach (Equation 3), the hazard index (HI) calculation (Equation 4) and the calculation of the increase in the probability of liver cancer attributable to AFM1 intake to assess the risk of the Hungarian population. For the MoE method, the $BMDL_{10}$ value of 0.4 μg/kg/ day for AFB1 was taken into account by a multiplication factor of ten (4 μg/kg/day) because AFM1 is ten times less potent carcinogen than AFB1.

$$MoE = \frac{BMDL_{10}}{EDI} \qquad \textit{(Equation 3)}$$

To calculate the hazard index, I used the safe dose recommended by Kuiper-Goodmann (0.2 ng/kg/day), which is the quotient of the tumor-causing dose in 50% of the animals and a safety factor of 50,000. Calculation of the aflatoxin hazard index proposed by Kuiper-Goodman:

$$HI = \frac{EDI\ (ng\ kgbw^{-1}day^{-1})}{0{,}2\ ng\ kgbw^{-1}day^{-1}} \qquad \textit{(Equation 4)}$$

The incidence of hepatocellular carcinoma associated with aflatoxin exposure was estimated by applying Equation 5, assuming hepatitis B prevalence of 0.7% in the Hungarian population:

$$R_{Hu} = [(P_{HBV+} \times HBV+) + (P_{HBV-} \times (1–HBV+))] \text{ x EDI} \qquad \textit{(Equation 5),}$$

where R_{Hu} is the risk of liver cancer incidence in the Hungarian population, HBV+ is the prevalence of chronic hepatitis B in the Hungarian population and P_{HBV+} is the probability of developing liver cancer in this part of the population, and P_{HBV-} is the probability of developing liver cancer in the rest of the population. I also performed the calculations for both optimistic and

pessimistic scenarios, for the latter (CI95 R_{Hu}), in case of 1 ng/kg bw intake of AFB1 per day taking into account the upper limit of the 95% confidence interval for the probability of developing liver cancer.

$$Average\ R_{Hu} = [0.027 \times 0.007) + (0.002 \times 0.993)] \times EDI \quad (Equation\ 6)$$

$$CI95\ R_{Hu} = [(0.056 \times 0.007) + (0.005 \times 0.993)] \times EDI \quad (Equation\ 7)$$

2.3 Probabilistic methods

For the probabilistic estimations, I fitted different distributions to the analytical results above the LOQ with the GAMLSS and GAMLSS.dist packages of the R statistical software using the maximum likelihood estimation. Then I selected the distribution that gives the optimal fit by the parameters describing the goodness of fit (AIC - Akaike's Information Criterion, BIC - Bayesian Information Criterion and Global Deviance). For AIC, BIC, and Global Deviance as well, the distribution with the smallest value should be considered the best fit.

The goodness of the fits was also evaluated by visual comparison of the histograms made from the data and the obtained distribution, as well as by examining the normality of the differences and using the Q-Q plot. Both the residual statistics and the QQ plot examine the differences between the original and the fitted data, then compare the data set of the residuals to a standard normal distribution and use a correlation coefficient to examine data point by data point, how much they deviate from it.

The two best-fit distributions were the two-parameter lognormal and the four-parameter Box-Cox t (BCT), which are suitable for modeling aflatoxins-like positively or negatively skewed, slowly decaying data, with continuous distribution.

The selected distributions were then fitted to the entire AFM1 data set.

After that, I continued to work with two types of probabilistic methods:

In the case of the first, Probabilistic Method I (Prob. I.), I generated 200,000 - 200,000 values from both the average daily consumption data calculated per consumer and the AFM1 concentrations measured in milk samples by re-sampling (20 x 10,000 iterations), from which I calculated 200,000 exposure values. The values of the concentration data below the LOQ were taken into account with the value of the imputed (generated) data equal to the number of

measurement results. Imputation was performed using the descriptive parameters of the lognormal distribution fitted to the concentration values and the selected most typical LOQ range (> 5 ng/kg). The relative and cumulative frequencies of the exposure data calculated for different consumer age groups characterize the expected exposure probability values.

For the other - Probabilistic II. method (Prob. II.) - I used the two-dimensional Monte Carlo model. The Monte Carlo simulation generates samples from the probability distribution fitted to the data by random sampling. The advantage of the Monte Carlo method is that not only original values between the minimum and maximum of the analytical data are selected as in the simple random resampling procedure, but the full spectrum of values below the distribution curve is used for the calculations. Values at both edges of the distribution play a particularly important role, which may play an important role in modeling.

The Monte Carlo model works with an external and an internal simulation loop. In the inner loop, the model performs the exposure calculation several times, randomly sampling consumption and concentration data, calculating different percentiles of exposure from each iteration (this is the variability of exposure). The sum of these exposure calculations constitutes an iteration of the outer loop and results in an estimate of the distribution of exposures. The outer loop also runs several times, and since repeated iterations will necessarily result in different percentile values due to random sampling, their distribution is characterized by uncertainty in the estimation.

So, in summary, the inner loop simulates the expected variability in daily exposures and the outer loop simulates the estimation uncertainty. At the end of the calculation series, the model characterizes the expected exposure of the population using the cumulative frequency distribution graph as well as percentile values. The graph shows 50% and 95% uncertainty intervals of the 2.5^{th}, median and 97.5^{th} percentile estimates, over the full spectrum of the estimated exposure.

2.4 The KNIME software

The calculations were performed using KNIME (Konstanz Information Miner) software. It is a free, open-source data analysis software. With KNIME, data analysis building blocks ("nodes") can be connected to create complete workflows, each node performing a computational operation or data function.

KNIME is suitable for storing, in a modular form, the steps to be performed with the data in a single file, as well as the data itself. Both the data entered and the steps to be performed with the data can be freely edited, nodes can be individually labeled, captions, additional information or even instructions can be added to workflows or certain parts of them.

The program has R and Python (programming language) integrations, so R and Python codes can be run within KNIME to perform computational tasks for which there is no built-in KNIME module. The developed exposure estimation methodologies can be made available to anyone, freely modified, optimized, easily adapted to other contaminant-matrix combinations, or expanded with additional modules or data sources.

3 RESULTS

3.1 Database of AFM1 processing factors for dairy products

AFM1 processing factors for dairy products as well as different cheeses (hard, semi-hard, soft and processed cheeses, fresh cheeses) were collected from the latest literature data.

In the 2009 and 2018-2020 food consumption surveys, I classified the AFM1-relevant food categories into the categories of processed dairy products and provided them with processing factors. The calculations were performed with the minimum, median and maximum values of the processing factors.

The database currently covers 85% of the AFM1-relevant food categories of the food consumption surveys (excluding butter, buttercream, condensed milk, milk powder, cream and ice cream). The table is used as a source database by the exposure estimation model, so the results of the estimates can be automatically further refined by re-running the calculations if the initial database is supplemented with additional data.

3.2 Changes in the frequency of milk and milk products consumption based on the 2009 and 2018-2020 surveys

I compared the changes in the frequencies of consumption of milk and different dairy products using the milk and dairy product consumption days of the 2009 and 2018-2020 surveys. I compared the number of consumption days of different foods to the total consumption days of the given survey. The method therefore characterizes the frequency of consumption of different foods during the survey periods. Among the food categories, the consumption

frequency of milk and milk-based desserts increased by circa 20%. The frequency of cheese consumption shows an increase of 14%. The frequency of consumption of sour milk products (kefir, yogurt, sour cream), cream and flavored milks remained almost constant. The consumption frequency of condensed milk and milk powder has decreased significantly. Overall, the frequency of consumption of milk and milk products has increased slightly over the last 10 years.

Based on the change in consumption frequencies over 10 years, an increase in aflatoxin exposure could be expected, however, this effect was counterweighed by the change in consumption volumes. The average consumption in milk equivalent, calculated with the median value of processing factors, was 310.7 g/day in 2009, this value decreased to 295.3 g/day in the 2018-2020 period.

3.3 Results of exposure assessments

In the following, I compare the results of the exposure estimates from different aspects, which were calculated by different methods. Differences in results were analyzed by analysis of variance (ANOVA) at a significance level of 0.05.

3.3.1 Comparison of results calculated taking into account the minimum-median-maximum values of the processing factors

To compare exposure values, I chose the average and 97.5th percentile results of the 2018-2020 (EU MENU) survey calculated by the deterministic method. Consideration of the minimum-median-maximum values of the processing factors did not significantly affect the results. There was a significant difference only in the mean values of the age group of the toddlers, in the case of the minimum and median factors, therefore hereinafter I use the values calculated with the median of the processing factors to present the different exposure estimation results.

3.3.2 Comparison of exposure of consumer age groups

The next evaluation criterion was to compare the exposure of different consumer age groups, based on the 2.5th percentile, mean, median, and 97.5th percentile estimated daily intake (EDI - ng/kg bw/day). The calculations were made based on data from the 2009 food consumption survey, using a deterministic method, taking into account the median of processing factors and the mean of AFM1 concentration data.

Taking into account the 95% range of estimation uncertainty, the exposure of toddlers is in the range of 0.03-0.55 ng/kg bw/day and it can be characterized by a mean value of 0.26 ng/kg bw/day (standard deviation 0.14 ng/kg bw/day) and a median value of 0.24 ng/kg bw/day. The exposure of children ranges from 0.04 to 0.34 ng/kg bw/day, with a mean value of 0.15 ng/kg bw/day (standard deviation 0.09 ng/kg bw/day) and a median value of 0.13 ng/kg bw/day. The exposure of adolescents, adults and the elderly is much lower, ranging from ≤ 0.01 to 0.17 ng/kg bw/day, with an average value of 0.04-0.06 ng/kg bw/day and a median value of 0.03-0.06 ng/kg bw/day. The difference between the age groups can be considered significant (p-value = 0.003).

Thus, the highest exposure values can be observed at the youngest and the lowest exposure values at the oldest age groups. However, the relationship is not directly between age and intake, but between the average body weight observed in different age groups (typically increasing by age) and intake, as exposure values are given per kilogram of body weight.

3.3.3 Comparison of exposure results for 2009 and 2018-2020

To examine the differences between the 2009 and 2018-2020 exposure estimates, I chose the results of the deterministic and probabilistic I. (random return) methods. In both cases, I present the results calculated with the median of processing factors. For the deterministic estimation I calculated with the average of the AFM1 concentration data, in the case of the probabilistic method, I calculated with the data generated by the lognormal distribution fitted to the AFM1 concentrations.

Both by comparing average or 97.5th percentile estimated daily intakes, the exposure in each age group has been found to be mostly constant over the past 10 years. The only noticeable difference is in the average values of the age group of toddlers and the 97.5th percentile values of the age group of children, however, the difference was found to be statistically significant only for the latter (p value = 0.04).

I obtained very similar results with the probabilistic method, although it resulted in higher results at the 97.5th percentile, the difference between the 2009 and 2018-2020 exposure data for each age group – with the exception of the 97.5th percentile values for the toddler age group (p value = 0.04) – is not significant. Although consumption rates for milk and dairy products have increased slightly over the past 10 years, this has not resulted in an increase in aflatoxin exposure because the decline in average consumption has counterweighed this change.

3.3.4 Comparison of results calculated by deterministic and probabilistic methods

In the following, I compare the results calculated by the deterministic, the probabilistic I. method, and the probabilistic II. method (2-dimensional Monte Carlo model) with Box-Cox t- (BCT) and lognormal (LogNorm) distributions. All calculations are based on the results of the 2018-2020 consumption survey. The median, mean and 97.5th percentile EDI values of the age group of toddlers and adults calculated with different exposure estimates were compared.

In case of both age groups, the median, mean, and 97.5th percentile estimates, calculated by the deterministic and 2D Monte Carlo methods resulted in very close values. There was no significant difference between the mean values by either method. In the case of probabilistic method, the median values were lower, the 97.5th percentile values were found to be higher than the others. It is likely that this is due to the nature of the random resampling method, as this type of estimation "amplifies" the two ends of the distribution.

Deterministic estimates usually result in a more conservative (pessimistic) estimate than probabilistic methods. In the present case, no higher deterministic exposure results can be observed, which is probably due to the fact that for AFM1 concentration data, not the mean AFM1 value was used (and results below LOQ not as LOQs), but the mean value of the generated dataset using lognormal distribution. This is thus in fact a semi-parametric estimate.

3.3.5 Comparison of results calculated with the 2D Monte Carlo method, Box-Cox t and lognormal AFM1 distribution

To compare the results calculated with the 2D Monte Carlo method with two different distributions (Box-Cox t and lognormal), I chose the data of the toddler and adult age groups of the 2018-2020 (EU MENU) survey.

Exposures calculated using the two different fitted distributions fall in roughly the same range, however, the values calculated with the lognormal distribution cover a wider range of uncertainty. Differences may be due to the nature of the probability distributions used as well as differences in the goodness of fit.

The mean and median values and associated 2.5th and 97.5th percentile values characterizing the exposure distribution of the age group of toddlers are summarized in numerical form in the tables below (Tables 1 and 2). They can

be used to compare the range of exposures calculated with the BCT and LogNorm distributions. Differences between mean and median values present the asymmetry of the estimate, indicating that the distribution is skewed toward small values.

Table 1: The mean, median and corresponding 2.5th and 97.5th percentile estimates of the age group of the toddlers calculated by BCT distribution using the 2D Monte Carlo method.

BCT Toddlers	mean	sd	min	1%	2.5%	25%	50%	75%	97.5%	99%	max
median	0.134	0.076	0.002	0.007	0.011	0.068	0.129	0.200	0.260	0.263	0.265
mean	0.161	0.091	0.003	0.008	0.013	0.082	0.156	0.240	0.313	0.316	0.319
2.5%	0.093	0.053	0.002	0.005	0.007	0.047	0.090	0.139	0.181	0.183	0.185
97.5%	0.400	0.227	0.007	0.002	0.031	0.203	0.387	0.597	0.777	0.785	0.793

Table 2: The mean, median and corresponding 2.5th and 97.5th percentile estimates of the age group of the toddlers calculated by LogNorm distribution using the 2D Monte Carlo method.

LogNorm Toddlers	mean	sd	min	1%	2.5%	25%	50%	75%	97.5%	99%	max
median	0.157	0.090	0.003	0.005	0.007	0.076	0.161	0.230	0.310	0.313	0.320
mean	0.187	0.107	0.003	0.006	0.009	0.090	0.191	0.274	0.369	0.372	0.381
2.5%	0.051	0.029	0.001	0.002	0.002	0.025	0.052	0.075	0.101	0.101	0.104
97.5%	0.488	0.281	0.008	0.016	0.023	0.235	0.500	0.715	0.965	0.972	0.996

The median, mean, and 97.5th percentile exposures values of toddlers modeled with a lognormal distribution are slightly higher, on the other hand the 2.5th percentile values, are lower than the same values calculated with BCT distribution.

As these two distributions characterized best the input data, overall it cannot be stated that one or the other results give more realistic estimate, but from the similarity of the results we can conclude that the real exposure values fall within the range covered by the two results.

3.4 Risk characterization of AFM1 intake

I used three evaluation methods for risk characterization. All methods are accepted in international practice, although the application of the hazard index and the margin of exposure (MoE) approach somewhat contradicts the fact that no safe tolerable daily intake can be established for genotoxic and carcinogenic compounds. Yet both methods establish a limit value compared to which some aflatoxin intakes are considered riskier and others less risky.

In any case, as the Estimated Daily Intake (EDI) alone does not provide sufficient information to judge whether the level of exposure can be considered low or high, these methods will help to assess the level of risk.

3.4.1 Risk characterization based on hazard index (HI) values

Dividing the result of the exposure estimates (EDI) by the safe dose, gives a dimensionless ratio. The extent of the risks is proportional to the results obtained and are considered to be of concern at values of 1 or higher. I used the results of deterministic estimates with consumption data from the 2018-2020 food consumption survey to compare HI values for age groups (Table 3).

Table 3: The comparison of average and 97.5th of percentile HI values derived from the EDI values calculated by deterministic estimation by age group. High risk values are indicated by bold numbers.

2018-2020	Toddlers	Children	Adolescents	Adults	Elderly
DET HI Average	**1.0**	0.7	0.3	0.2	0.2
DET HI P97.5	**2.8**	**1.9**	0.8	0.6	0.4

Note: DET: deterministic method, HI: Hazard Index

HI values calculated from the mean daily intake values and 97.5th percentile values indicate that the risk from exposure in the groups of adolescents, adults and elderly is not considered to be a concern. However, in the case of toddlers and children, the 97.5th percentile values (large consumers), the exposure is significantly higher than the level considered safe. One of the most important of the above results is the HI value of 1, which characterizes the average intake of toddlers, as it suggests that a significant proportion of this age group is exposed to AFM1 at a level of health concern.

The above evaluation was performed not only for the estimated daily intake values calculated with deterministic method, but also calculated by probabilistic methods. The comparison of HI values calculated from the results of different methods was performed on the example of toddlers and adults.

The results obtained confirm the conclusions drawn above. The mean exposure values for the age group of toddlers were found to be worrying in two of the three exposure estimation methods, and all of the 97.5th percentile

exposure values resulted in a hazard index above 1. With the probabilistic I method, the 97.5th percentile calculation for adults also resulted in an HI value above 1, however, it is known that this method gives a very conservative estimate for the upper percentiles.

Another option for characterizing the risk from AFM1 intake is the Margin of Exposure (MoE). For aflatoxins, the $BMDL_{10}$ value derived from AFB1-induced liver cancer studies in rats (400 ng/kg bw/day) may be used as a reference value, which can be used for AFM1 converted by a factor of ten. Results below 10,000 are of concern, MoEs of 10,000 or greater indicating little risk to public health.

The average and 97.5. percentile margin of exposure estimates (MoE) of the intake values calculated by deterministic method from the consumption data of the 2018-2020 food consumption survey were compared by age groups.

The results from the margin of exposure assessment provide us with a less worrying picture than the output values of the hazard index calculation. The limit of considerable risk (10,000) was reached only by the 97.5th percentile of the toddler age group and approached by 'large consumers' of children. For the other age groups, no significant risk can be identified with this risk characterization methodology.

Comparison of MoE results obtained from deterministic and probabilistic methods from the consumption data of the 2018-2020 food consumption survey leads to similar conclusions. The average intake values of the age group of toddlers and adults are not considered risky, while the 97.5th percentile results fell into the critical range with all exposure estimation methods.

A third method for risk characterization is to estimate the contribution of mean and high AFM1 intake in a given population to the incidence of hepatocellular carcinoma (HCCi), i.e., the incidence of new cases in a given population over a given period of time.

Exposure to aflatoxin increases the risk of developing HCC in the presence of chronic hepatitis B. As the prevalence of hepatitis B is low in Hungary (and in Europe in general), the aflatoxin-induced increase in HCCi does not show high values either. Although the numerical value of the estimated incidence of liver cancer proved to be very low, the relative values of the results in this case also show higher risk for toddlers (0.00036-0.0023 cases/100,000 persons/year) and children (0.00026-0.0017 cases /100,000 persons/year)

compared to other age groups (0.000074-0.00039 cases/100,000 persons/year).

3.5 An integrated risk assessment model developed in the KNIME framework

To perform the calculation steps of the presented results, I created an integrated risk assessment KNIME workflow suitable for the processing of consumption data, deterministic and probabilistic exposure estimates, as well as for the characterization of exposure based on the obtained results. The model was developed in two versions, separately optimized for the data structure of the two consumption surveys used. The program is currently designed for milk and dairy product data as well as AFM1 concentration data processing, but can be run or expanded to any food-contaminant combination with minor modifications.

The integrated risk assessment model includes four modules, the first for preparing food consumption data, the second for performing deterministic exposure assessment, the third for performing probabilistic exposure assessment I, and the fourth for probabilistic exposure assessment II. The output data of the food consumption data processing module (average daily consumption data) form the input to the exposure estimates.

3.5.1 Food Consumption Data Processing Module

The first step in the module containing the processing steps for food consumption data is to retrieve the required consumption data, the food categorization data (FoodEx) and the table containing the processing factors. The following are the steps for data cleansing and formatting, and linking the relevant data from the three input tables. EU MENU food consumption data is supplemented with FoodEx codes and translated into food names by the FoodEx coding table. The processing factors can be associated with the food consumption data using the food names, hereinafter the summary data table contains both of these data.

In the following steps, the module summarizes the individual dairy intake values for each consumption day, taking into account the minimum, median and maximum values of the processing factors, converted into milk equivalent in three parallel calculation series. The aggregate average values of consumption days per consumer, expressed in kg/kg bw, are transmitted by the module to the exposure estimation modules. The processing of food consumption data ends with three identical metanodes. The metanode first

separates the data by age group, then calculates descriptive statistics from the average consumption data per consumer for each age group, and then prints the data and descriptive statistics on separate worksheets of an Excel spreadsheet.

3.5.2 Deterministic exposure estimation module

The output data of the first module is read by the deterministic exposure estimation module. In parallel, it also retrieves AFM1 data (CSV Reader), generates descriptive statistics on it, and then performs exposure estimation using the calculated AFM1 average value (according to the OIM methodology). It passes on the calculated values into a metanode.

The last metanode of the deterministic exposure estimation module separates the exposure values by age group and calculates the risk characterization metrics (MoE, HI, HCCi values) for each age group. Finally, it exports both the exposure values and the risk characterization metrics by age group to separate worksheets of an Excel spreadsheet.

3.5.3 Probabilistic I exposure estimation module

The first step of the module reads the AFM1 data and the consumption data, the latter being separated by age group in the first sub-process. From the concentration values, the program then takes 200,000 random samples (or the amount specified for the sampling cycle and the number of sampling items) by random sampling, calculates descriptive statistics from the obtained database, relative and cumulative frequency values, and then plots the obtained frequencies.

In parallel, the module acts similarly with the consumption data and then performs the exposure estimation by multiplying the sampled concentration and consumption data. The next sub-process calculates the descriptive statistics of the obtained exposure values and their relative and cumulative frequency values, and then plots the results on an interactive relative and cumulative frequency diagram. The last steps of the module are to calculate the risk characterization metrics and then export the obtained results to an Excel file.

3.5.4 Probabilistic II. exposure estimation module

The last, probabilistic II. exposure estimation module consists of four sub-processes. The first sub-process reads the necessary data and separates the age

groups. The next step is to fit the examined distributions to the AFM1 data above the LOQ, as well as to check the goodness of the fit, the obtained histograms and Q-Q plots are exported by the module in image format.

In the next sub-process, the distributions selected during the second sub-process are fitted to the complete data set (including values below the LOQ), and the goodness of fit is re-examined. In the optimal case, a positive change can be observed in the parameters describing the goodness of fit.

The last sub-process of the probabilistic II. exposure estimation is to run the Monte Carlo simulation with the selected distributions, for the data of each consumer age group. The result of the simulation is a cumulative frequency graph characterizing the expected exposure of the studied populations and a table containing the percentile values characterizing the variability. The calculation steps, used during the probabilistic II. exposure estimation, were performed by using the statistical packages written for the R statistical software, integrated into the KNIME framework.

4 CONCLUSIONS AND RECOMMENDATIONS

In its 2020 risk assessment, EFSA drew attention to the concerns about exposure results for both aflatoxin B1 and aflatoxin M1, especially in young age groups. As the report did not contain detailed information on the exposure of Hungarian consumers, I considered it important to estimate the domestic aflatoxin exposure as accurately as possible using the appropriate methods.

Analysing the obtained results in the light of the European data and the results of the previous collaboration with Italy (Table 4), it can be concluded that the exposure pattern of Hungarian consumers, the relative exposure of age groups, follows the results calculated by EFSA and obtained from the Italian data. Based on the exposure data, the exposure of Hungarian consumers to AFM1 is lower than the average and 95th percentile AFM1 exposure values calculated by EFSA and the results for Italian consumers.

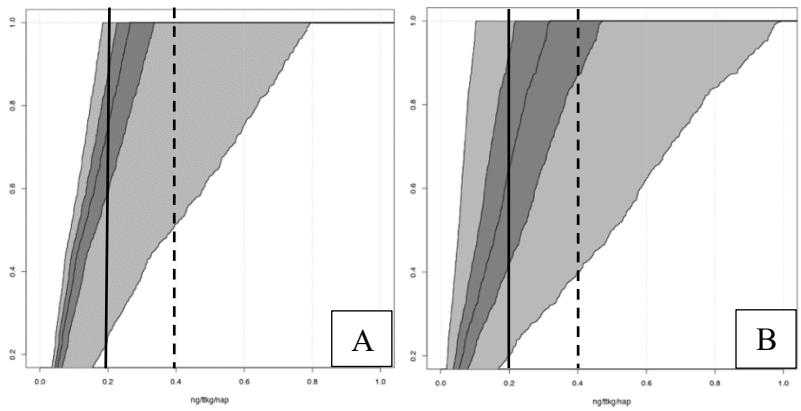
Table 4: Mean and 95th percentile exposure values (ng/kg bw/day) calculated by the deterministic method for 2018-2020, compared to the results of EFSA 2020 and AFM1 exposure calculations in Italy

	Infants	Toddlers	Children	Adolescents	Adults	Elderly
EFSA 2020 Average	0.69	0.86	0.43	0.19	0.1	0.1
IT 2018 Average	0.33	0.28	0.10	0,04	0.03	0.03
HU 2020 Average	?	0.19	0.14	0.06	0.04	0.03
EFSA 2020 P95	1.77	1.82	1.03	0.47	0.57	0.28
IT 2018 P95	0.88	0.63	0.19	0.07	0.07	0.05
HU 2020 P95	?	0.44	0.33	0.13	0.10	0.07

Note: IT: Italian result, HU: Hungarian results, P95: 95th percentile exposure values

It is important to note that for both benchmark surveys, the age group of infants was also included in the risk assessment (this age group was not examined in the 2018-2020 survey) and their exposure values approximate, in some cases exceed, those of the toddler age group. This leads to a conclusion that infants in Hungary also belong to the vulnerable population groups, however, we are not able to support this with data resulting in a reliable estimate.

Based on current data for the age group of toddlers and those who consume large amounts of milk and dairy products among young children, it can be concluded with high certainty that they are at health risk in terms of AFM1 intake. I present at Figure 1, using a combination of the previously presented exposure results calculated using the Monte Carlo method and the HI and MoE metrics used to characterize the risk, what proportion of these two age groups are considered at risk.



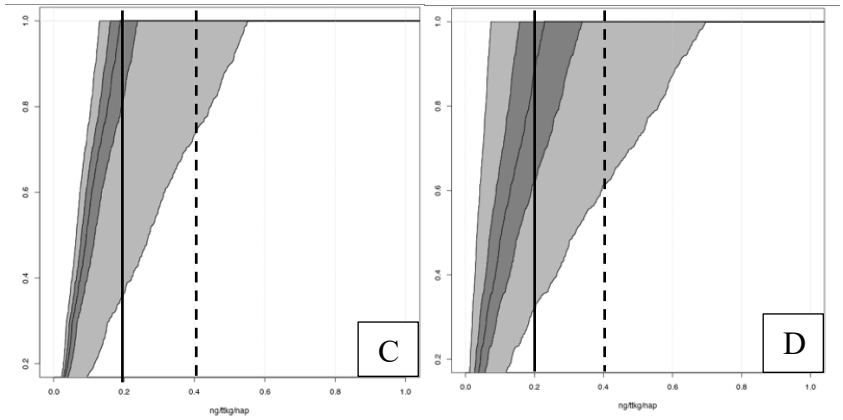


Figure 1: Cumulative frequency distributions of AFM1 exposure estimated by the 2D Monte Carlo method based on two distributions (BCT on the left, LogNorm on the right) of the age groups of toddlers (A, B) and young children (C, D) in the 2018-2020 consumption surveys indicating the HI (solid line) and MoE (dashed line) thresholds.

In Figure 1, the cumulative frequency distributions of the age group of toddlers and young children calculated with the Box-Cox t- and lognormal AFM1 distribution are supplemented with the 0.2 ng/kg bw/day (left vertical line) used for the HI calculation and the 10,000 MoE values resulting in 0.4 ng/kg bw/day (dashed line) risk thresholds. The two diagrams provide a good indication of the distribution of exposures in the toddler and young age groups and also provide an opportunity to assess the degree of risk.

Given that the toxicity of aflatoxins poses primarily a health risk to the youngest population groups, special attention should be paid to reducing and minimizing their exposure. However, it is emphasized that the presence of carcinogens should be kept to a minimum in all age groups.

As a sufficient number of test results were provided by aflatoxin M1 concentrations measured in milk, I estimated the exposure from the intake of milk and dairy products using derived data. The partial results already show that certain age groups of the Hungarian population are exposed to a higher aflatoxin level than the safe dose and this can be considered as a kind of indicator. Therefore it is worthwhile to extend the exposure estimation in the future to AFB1, which is ten times more toxic, by analyzing concentration values and consumption data for foods relevant to AFB1, and cumulative risk assessment of different aflatoxins can also provide informative results.

However, this requires adequate quantity and quality of data on other foods relevant to aflatoxins (e.g. processed cereal-based products, dried fruits, spices, etc.). Data in the literature show that total diet studies provide more accurate information for estimating exposure than the use of monitoring results, therefore, if possible, I recommend cunducting a national, representative total diet study, which – using appropriate sensitivity (HPLC-MS/MS) analytical methods – would allow the combined assessment of the levels of several contaminants of food safety concern, including all mycotoxins in food.

According to the annual monitoring results, milks with a contamination exceeding 10-15 times the maximum tolerable level are also marketed. A particularly vulnerable group is those who regularly consume milk from the same source where the animals are fed aflatoxin-contaminated feed.

Based on the results, I consider it justified to introduce measures to protect the youngest age groups and reduce the amount of aflatoxins entering the milk and milk products value chain.

As there is no large-scale process that can perfectly eliminate aflatoxins from the food chain, prevention remains paramount. Great emphasis should be given on the overall control of *Aspergillus* infection, and the use of agricultural, storage and processing technologies that inhibit the growth of molds and reduce aflatoxin levels should be promoted.

This is a complex task that requires the involvement of all actors in the food chain, starting with the application of good agricultural practices and the proper preparation and management of arable land. This is followed by a series of measures taken during the selection of hybrids resistant to mold, harvesting, transport and storage of crops, which can prevent the growth of molds (setting appropriate temperature and humidity levels, sorting, peeling, physical treatment of crops). Last but not least, the appropriate storage, handling and control of the aflatoxin content of cereals, silage or other processed feed preparations intended for animal feed and their physical, chemical or biological detoxification if necessary.

The success of prevention and the adequacy of milk shipments can also be checked at the level of dairy farms and dairies. A sampling plan and early warning system for the detection of aflatoxin M1 in raw milk can be used to effectively predict an increase in the level of contamination, using the 20 ng/kg action limit already proven in practice in Italy. According to the indication, the dairy farm can prevent the legally tolerable maximum (50

ng/kg) AFM1 concentration from being reached in a way that is appropriate to local conditions, for example through feed interventions. This would reduce the use of large quantities of aflatoxin-contaminated milk in primary and secondary milk processing and consequently reduce the exposure of consumers.

It should also be noted that ELISA kits set to indicate a concentration of 50 ng/kg AFM1 may still consider adequate a batch of milk contaminated with ≤ 65-70 ng/kg AFM1 in 50% of cases due to detection uncertainty.

For the group of infants and young children most exposed to AFM1 and at the same time most vulnerable, but also in order to protect the health of the whole population, I propose to amend the control of milk processing plants. I recommend that they are obliged to set the ELISA kit to indicate ≥ 20 ng/kg concentration, notify the dairy farm and NÉBIH if AFM1 contamination is ≥ 20 ng/kg in the milk delivered from a given farm, and thereafter to monitor the effectiveness of the measure taken to reduce the contamination by daily monitoring the contamination of the milk delivered from the same dairy farm.

It is also appropriate to regularly monitor consumer exposure according to the developed methodology using the latest aflatoxin analytical results. In addition to informing consumers and promoting a diverse diet of impeccable quality and the appropriate frequency of official controls, it is necessary to make business operators along the food chain economically interested in reducing the contamination of food and feed.

5 NEW SCIENTIFIC RESULTS

1. I applied mathematical equation for the calculation of the intake of aflatoxin M1 from milk taking also into account the processing factors of relevant milk based foods.

2. I developed an integrated model for the deterministic and probabilistic exposure assessment of the Hungarian population, including processing factors, as well as a food consumption database in KNIME framework. After the preparation of concentration data, the application can be used to fit distributions, for goodness of fit analysis, and by using the parameters characterizing the distribution, to perform a 2D Monte Carlo simulation and estimate the exposure from the intake of any chemical contaminant.

3. I estimated the aflatoxin M1 exposure of the Hungarian population from the consumption of milk and dairy products by using probabilistic and deterministic methods. My results indicate that the chronic AFM1 intake of adolescents, adults and the elderly does not reach the level of health risk, but the AFM1 intake of the age group of toddlers consuming average or large amounts, as well as the large eater children, exceeds the level of health risk.

Projects and funds related to the dissertation

The research related to my dissertation was carried out within the framework of the project 2018-1.2.1-NKP-2018-00002, with the support provided by the National Research Development and Innovation Fund, in the financing of the 2018-1.2.1-NKP application program.

6 PUBLICATIONS RELATED TO THE DISSERTATION

Articles published in peer-reviewed scientific journals (with IF)

Andrea Serraino, Paolo Bonilauri, **Kata Kerekes**, Zsuzsa Farkas, Federica Giacometti, Alessandra Canever, Angelo Vittorio Zambrini, Árpád Ambrus (2019) Occurrence of Aflatoxin M1 in Raw Milk Marketed in Italy: Exposure Assessment and Risk Characterization, Frontiers in Microbiology, 10:2516, https://doi.org/10.3389/fmicb.2019.02516 / IF: 4.235

Kata Kerekes, Paolo Bonilauri, Andrea Serraino, Federica Giacometti, Silvia Piva, Angelo Zambrini, Alessandra Canever, Zsuzsa Farkas & Árpád Ambrus (2016) An effective self-control strategy for the reduction of aflatoxin M1 content in milk and to decrease the exposure of consumers, Food Additives & Contaminants: Part A, 33:12, 1840-1849 /IF: 2.34

Articles published in scientific journals without IF

Ambrus Árpád, Szenczi-Cseh Júlia, Griff Tamás, **Kerekes Kata**, Miklós Gabriella, Vásárhelyi Adrienn, Szigeti Tamás János (2020) Élelmiszereink mikotoxin és növényvédőszer-maradék szennyezettségének élelmiszerbiztonsági megítélése, 1. rész. Növényvédőszer-maradékok; Élelmiszervizsgálati Közlemények, 66:1, 2772-2789

Kerekes Kata, Csorba Szilveszter, Ambrus Árpád (2021) A magyar fogyasztók krónikus aflatoxin M1 expozíciója; Élelmiszervizsgálati Közlemények, (Megjelenés alatt)

National and international conference publications, lectures

Kerekes Kata, Dr. Farkas Zsuzsa, Dr. Ambrus Árpád. Korai előrejelzési rendszer a tej aflatoxin M_1 tartalmának detektálására. Szent István Egyetem *Mikotoxin-fórum.* Budapest, 2018. október 2.