ImproRisk model, an open access risk assessment tool

Demetris Kafouris, Aspasia Sarandi, Georgios Stavroulakis, Christopher Papachrysostomou

State General Laboratory, Ministry of Health, 44 Kimonos St, 1451 Nicosia, CYPRUS

www.improrisk.com

Correspondence: dkafouris@sgl.moh.gov.cy

Abstract

ImproRisk is an open-access Risk Assessment model, developed in Cyprus, which contributes to the harmonization, standardization and transparency of the Dietary Exposure Assessment methodologies across Europe. It is a simple and transparent tool, built in R, for conducting accurate dietary exposure assessments to chemical substances. Aiming to be established as a standardized Risk Assessment tool at European level, the model has been designed to accept occurrence and consumption data, coded according to the EFSA's food classification and description system version 2 (FoodEx2). ImproRisk users are able to estimate dietary exposure to a chemical substance under study at any level of the EFSA's FoodEx2, considering the FoodEx2 base term and the Process facet (F28). Dietary exposure is estimated at individual level for different population groups and exposure estimates are stratified by different demographic characteristics. ImproRisk is considered as a useful tool both for risk assessors and risk managers, as it facilitates decision making, by providing information regarding the percentage of the population exceeding the relevant health-based guidance value or reference points for a specific chemical substance.

Keywords: ImproRisk, FoodEx2, Dietary exposure assessment,

Table of Contents

1. Introduction

Dietary exposure assessment is crucial, especially in the context of public health, nutrition research, and food safety. It plays a vital role in evaluating potential health risks associated with food contaminants, additives, or specific nutritional components. By quantifying exposure levels, researchers can assess the likelihood of adverse health effects and establish safe exposure limits or regulatory guidelines. Understanding what people eat is essential for monitoring population health. Dietary assessment helps identify trends in food consumption, nutrient intake, and dietary patterns, which are directly linked to the prevalence of various chronic diseases such as obesity, diabetes, cardiovascular diseases, and certain cancers. Monitoring dietary exposure allows identifying potential food safety hazards such as microbial pathogens, toxins, pesticides, and chemical contaminants. Such information is crucial for implementing food safety regulations, conducting risk assessments, and ensuring the safety of the food supply chain. Governments, health organizations, and stakeholders rely on accurate dietary data to formulate strategies aimed at improving dietary habits and promoting healthy eating behaviors. Assessing dietary exposure assists in educating individuals and communities about healthy eating habits, portion sizes, balanced diets, and the importance of diverse food choices. It empowers people to make informed decisions about their diets, leading to improved overall health and well-being. In the context of global health challenges such as malnutrition, food insecurity, and diet-related diseases, accurate dietary exposure assessment is essential for identifying vulnerable populations, implementing targeted interventions, and evaluating the effectiveness of nutrition programs on a global scale. Overall, dietary exposure assessment is a multidisciplinary endeavor that intersects with nutrition science, public health, food safety, policy development, and consumer education. Its importance lies in its ability to provide data-driven insights that contribute to promoting healthier eating behaviors, preventing nutritional deficiencies and diseases, and ensuring the safety and sustainability of food systems.

The aim of the development of a simple and accurate deterministic exposure assessment model was to support risk assessors from National Food Safety Authorities, Risk Assessment Institutions and Academia to perform chronic dietary exposure assessment at individual level using their national food consumption data.

2. Main features and benefits of the ImproRisk model

ImproRisk model was developed to estimate chronic dietary exposure to different chemical substances present in food. It allows the estimation of mean exposure as well as the exposure in several percentiles (P25, P75 and P95) in different population groups, across gender and different geographical regions within a country. Users of the ImproRisk Model have the ability to conduct chronic dietary exposure assessment to chemical substances at FoodEx2, which allows a more accurate linkage between occurrence and food consumption data due to the availability of detailed food groups. Additionally, exposure assessment can be conducted by exploiting FoodEx2 processing facets (F28). Moreover, users can upload food consumption datasets that include subjects with different demographic characteristics (age groups, areas of living), and further explore the uploaded data. ImproRisk can also estimate weighted exposure by applying weighting coefficients to adjust the sample for non-representativeness within the population. What is more, it has the capacity to produce customized table and graph reports as well as to export all data in excel worksheets for further processing and evaluation. Furthermore, users can "explore" exposure at individual level and at the consumption occasion level, and generate automatically the final exposure assessment report.

The model is hosted online, providing easy access to users. The model utilizes the computational engine of R, a statistical analysis software coupled with a programming language. The R language facilitates future development and upgrades. It is interactive, since users can interfere and configure the tables/graphs. Users can work on their data, save the progress and resume at a later stage with their latest uploaded data. Validation checks of input datasets for both occurrence and food consumption data are performed during uploading. The model is linked to the EFSA's Chemical Hazards Database OpenFoodTox database for easy access to essential info for Risk Assessment e.g. reference values, reference points etc. (https://www.efsa.europa.eu/en/data-report/chemical-hazards-database-openfoodtox). ImproRisk also includes a module namely 'Simple Risk Assessment', that is compatible to accept food consumption data from the open access EFSA's Comprehensive Food Consumption Database (https://www.efsa.europa.eu/en/data-report/food-consumption-data).

3. Access to ImproRisk

ImproRisk is an open access tool. A two-step process is needed to gain access:

i. **Free registration on the ImproRisk website.**
This will be reviewed by SGL and, upon approval, users become eligible to create an account with their email address on the ImproRisk online application.

ii. Authorization of the new user's email account:
An email will be sent to the users stating that they can use the registered email address to create an account on the ImproRisk application. The email will contain the link to the application.

A more detailed process is described below:

- Visit ImproRisk website at http://www.improrisk.com/conditions and register by filling the relevant form with all the necessary information.
- Notification of registration will be sent to your personal mailbox.
- Upon approval by the SGL Risk Assessment Unit, the link to the ImproRisk model is sent to the provided email address. This is to notify you that an account can be created on the ImproRisk application using your registered email address.
- Go to https://sglcy.shinyapps.io/improrisk/, fill in your email address and choose the option "First time user? Register here!".
- Create an account where your username will be your email address and a password of your choice.

4. Data inputs

4.1 Occurrence data

Occurrence records must be codified according to the EFSA's food classification and description system, FoodEx2 (EFSA, 2015). Occurrence data must be codified at the appropriate FoodEx2 Level, of the exposure hierarchy, and aggregated according to their base term and processing facet (F28), as presented in Table 1.

ImproRisk users are required to input to the model the occurrence values for the chemical substance of interest either by uploading their own file in excel format with aggregated occurrence records or by creating their own aggregated occurrence data directly in the model. The substitution method is applied for the aggregation of left censored (LC) occurrence data (EFSA, 2010). According to this approach, non-detects must be replaced with the value of zero at the Lower Bound (LB) scenario, half the Limit of Detection (LOD/2) or the Limit of Quantification (LOQ/2) at the Middle Bound (MB) scenario and the LOD or the LOQ at the Upper Bound (UB) scenario. ImproRisk assumes that all the occurrence data are expressed in mg/kg of food.

The uploaded aggregated occurrence dataset must conform to specific data validation rules with regards to column naming and content. Valid FoodEx2 code and no duplicate values are mandatory. Additionally, mean occurrence at LB, MB and UB scenarios should be positive values (or zero). The model automatically checks and validates the dataset upon uploading, and reports regarding Column Content, Missing Values and Invalid Entries. Table 1 shows the mandatory data fields for the occurrence.

Table 1. Mandatory data fields for occurrence data input

termCode	LB_mean	MB_mean	UB_mean
A0DPP	0,148	0,148	0,148
A00ZT#F28.A07GR	0,2989	0,2989	0,2989
A00ZT#F28.A07GL	0,045	0,045	0,045
A00ZT#F28.A07GY	0,103	0,103	0,103
A03PV	0,0287	0,0298	0,031
A03QX	0,0311	0,0324	0,0336
A03RC	0,024	0,025	0,026
A03QZ	0,0138	0,0154	0,0169
A03RA	0,0772	0,0777	0,0781
A03RF	0,024	0,025	0,026
A03VA	0,4956	0,4958	0,4961
A03VB	0,1038	0,1042	0,1045

4.2 Food consumption data

In general, the ImproRisk model was designed to facilitate the exploitation of food consumption data collected during the implementation of the EU Menu project by the European countries, according to the Guidance on the EU Menu methodology (EFSA, 2014), in order to be used for conducting dietary exposure assessments. Only those individuals with at least two consumption days reported are taken into account for the exposure calculations.

All consumption records intended to be used as inputs to the model must also be codified according to EFSA FoodEx2 (EFSA, 2015). The ImproRisk model allows the use of any FoodEx2 term at any of the seven Levels of the Exposure Hierarchy. Facet descriptors regarding processing (F28) and ingredient (F04), in addition to the FoodEx2 base term, are currently represented in the ImproRisk model. The model is capable to recognize the processing and the ingredient facet descriptors that are present in the food consumption dataset, which is uploaded by the user. This function enables the exploration of the specific facets within the uploaded dataset. Users will be able to estimate exposure, considering the FoodEx2 base term and the processing facet (F28). The model allows the users to include demographic data for the population under study, within the uploaded consumption dataset, such as age, gender and area of living within the country. Exposure estimates are also stratified by these demographic characteristics. Table 2 shows the mandatory data fields for the consumption.

Table 2. Mandatory data fields for consumption data input.

SERIAL	SUBJECTID	DAY	AMOUNTFOOD	AMOUNTFCOOKED	FOODEXCODE	GENDER	AGE	WEIGHT	AREA	POP_CLASS	WCOEFF
300177	20001	1	40		A00DD#F19.A07PR$F28.A07MS	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300178	20001	1	100		A02LZ#F07.A06ZA$F10.A0CQD	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300179	20001	1	70		A004Y#F19.A07PR$F22.A07SZ	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300180	20001	1	20		A023V#F19.A07PR	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300181	20001	1	15		A065R#F19.A07PR$F27.A02LV	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300182	20001	1	250		A03DL#F19.A07PR	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300183	20001	1	70		A00BG#F19.A07PM$F22.A07SS	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300184	20001	1	45	120	A013N#F19.A07PR$F28.A07GL	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300185	20001	1	5		A036P#F19.A07PR$F22.A07SH	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300186	20001	1	25		A0DMX#F19.A07PR$F28.A07JG$F28.A07KY	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300187	20001	1	20		A004Y#F19.A07PR$F22.A07SZ	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300188	20001	1	45		A034J#F19.A07PR	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300189	20001	1	75		A01CD	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300190	20001	1	80		A00QH#F20.A07QF$F28.A07JG	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300191	20001	1	2		A03AH	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300192	20001	1	2		A042P#F19.A07PR	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300193	20001	1	269	180	A01SP#F02.A07XQ$F19.A07PR	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300196	20001	1	15		A00KX#F19.A07PR$F28.A07JG	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265
300197	20001	2	40		A00DD#F19.A07PR$F28.A07MS	FEMALE	10,44764	31,3	Lefkosia	Adolescents	265

4.3 Weight coefficients

The ImproRisk model accommodates weight coefficients to adjust the sample for non-representativeness within the population. Exposure statistics, such as mean and percentiles, are adjusted to reflect the population from which the sample was taken from. The model applies frequency weights which are integer values above 1 (≥1), that indicates the number of persons in the population that a subject in a survey sample represents. If, for example, the weight coefficient for a subject in the sample is equal to 500, that subject represents 500 persons in the population of the corresponding class. Typically, weight coefficients are calculated using (recent) census data of the population. Weight coefficients need to be calculated and incorporated in the consumption dataset prior uploading the relevant dataset by the users. In case there are no weight coefficients assigned, the weight for each subject has to be set to "1".

4.4 Substance

Users can set the required information for the chemical substance for which the exposure estimate will be calculated. The name of the substance should be provided and it will be appeared in Graph and Table captions and the exported report. A choice of one substance category, such as Additive, Pesticide, Veterinary Drug Residue, Contaminant and Genotoxic-Carcinogen is available, and will appear in the exported report. Users can choose a reference value such as Acceptable Daily Intake (ADI), Tolerable Daily Intake (TDI), Tolerable Weekly Intake (TWI), Provisional Maximum Tolerable Daily Intake (PMTDI), Provisional Tolerable Weekly Intake (PTWI), if applicable. Alternatively, a reference point, Benchmark Dose Level (BMDL) and No Observed Adverse Effect Level (NOAEL), can be selected. This will have an effect on the calculations as choosing a Reference Value or a Reference Point, will define the exposure as μg/kg b.w. per day or μg/kg b.w. per week, and the calculations will be based on a per day or per week basis, respectively. A numeric reference value in µg/kg b.w. per day or per week should be entered, since it will have an effect on the calculations with regards to the MoE in case of a BMDL or NOAEL (Reference Points), or the rate of exposure above a threshold, for the Reference Values.

5. Exposure assessment methodology

The model considers that each subject at each consumption occasion is exposed in to the chemical substance of interest, thus it combines the mean occurrence values at LB, MB and UB scenario with consumption data, at individual level, for the estimation of the subject's individual exposure. The estimated individual exposure can later be used to cross tabulate by other demographic characteristics such as gender, population class and area of living.

The estimated daily or weekly dietary exposure to a chemical substance is calculated according to the equation (1):

$$\text{Exposure } \left({}^{\mu g}\!/_{\text{kg b.w.}}\right) = \frac{\text{Consuption (g of food) x Occurence } \left({}^{mg}\!/_{\text{kg of food}}\right)}{\text{Body weigh t (kg)}} \quad (1)$$

The body weight is the subject's body weight. The outcome of the equation is the exposure (in µg/Kg b.w.) at each and every food consumption instance. The calculated exposure for each instance is summed for all days and for every food category at the respective FoodEx2 Level. In particular, the estimated daily exposure is calculated for each individual in the dietary survey, by taking the average exposure throughout the number of consumption days of the survey. In this way, the exposure rate is calculated for each individual at an "Optimistic" (LB), an "Average" (MB) and a "Pessimistic" (UB) scenario.

5.1 Occurrence and Consumption matching

ImproRisk combines the consumption and occurrence datasets using a level-up approach matching in the FoodEx2 exposure hierarchy. More specific, a food item in the consumption dataset is matched either to the same food in the occurrence dataset, or a parent food item (higher level), up to a maximum of two levels in the hierarchy. For example, consumption of lamb fresh meat, which is at Level 5 of the FoodEx2 hierarchy, has the base term code [A01RK]. In order to estimate the exposure for this food item, ImproRisk will perform the calculation by searching an occurrence value of the food item 'Lamb fresh meat' [A01RK], in the occurrence dataset. If found, it will use the relevant occurrence value to calculate the exposure. In the case where it is not found, it will search "upwards" in the hierarchy, meaning one level higher at a time, for occurrence in a parent food item. In this case, it will search for an occurrence value of "Sheep Fresh Meat" [A01RH] (Level 4). If concentration value at this level is available, it will be used for the exposure calculation. In the case that the information is not found at Level 4, it will continue searching upwards at Level 3 (Mammals Meat). Finally, if concentration value is still not available at Level 3, the process stops, and the calculated exposure is zero.

5.2 Estimated Dietary Exposure

Users can view the results regarding the estimated chronic dietary exposure to a specific chemical substance. Specifically, they can view the estimated overall exposure for the population, the estimated exposure by a demographic variable, as well as the estimated exposure by two demographic variables. One of the main features of the ImproRisk model is that it offers the possibility to the users to explore the results obtained regarding the estimated exposure for each subject and the calculations performed to derive the exposure.

The estimated overall exposure for the population includes descriptive statistics of the exposure as well as Probability and Cumulative distribution of exposure curves (Table 3, Figure 1). When assessing dietary exposure to chemical substances with an associated Reference Point (i.e. BMDL, NOAEL), the model automatically calculates the mean and P95 Margin of Exposure (MoE) values, by dividing the BMDL value with the estimated mean and P95 exposure, respectively.

Table 3. Statistics of the total population exposure to lead (Kafouris et. 2024).

	LB (μg/Kg b.w. per day)	**MB (μg/Kg b.w. per day)**	**UB (μg/Kg b.w. per day)**
Min	0.02	0.028	0.035
Max	1.909	1.931	1.953
Mean	0.136	0.17	0.205
95% C.I.	(0.126 - 0.146)	(0.16 - 0.181)	(0.194 - 0.216)
SD	0.127	0.14	0.155
P25	0.078	0.103	0.124
Median	0.109	0.138	0.169
P75	0.153	0.196	0.238
P95	0.304	0.371	0.44
MoE (Mean)	7.366	5.871	4.874
MoE (P95)	3.292	2.697	2.273

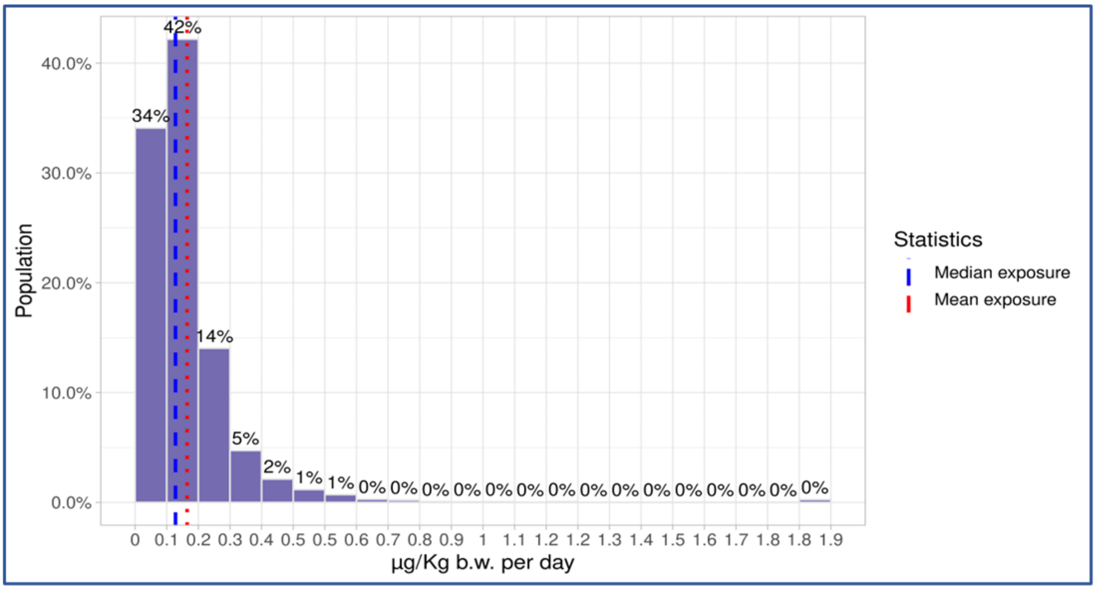

Figure 1. Probability distribution of exposure at MB scenario.

Users can select a demographic variable, to view the estimated exposure results per variable (i.e. population class, area of living and gender). Exposure statistics across the categories of the variable are presented in tables and illustrated in Bar plots of the mean exposure, and Probability/ Cumulative distribution of exposure curves across the categories (Table 4, Figures 2-4). Additionally, a statistical hypothesis testing is, conducted across the different groups of the demographic variables for the determination of statistically significant differences. For demographic variables consisting of two groups (e.g. gender), t-test is applied for statistical hypothesis testing in the estimated dietary exposure while for demographic variables consisting of more than two groups (e.g. population class or area of living), statistical hypothesis testing is conducted via Analysis of Variance (ANOVA) test (Table 4).

Finally, supports the combination of two demographic variables to produce a cross table and a cross plot of the average exposure across the levels of the two chosen demographic variables (Figure 5).

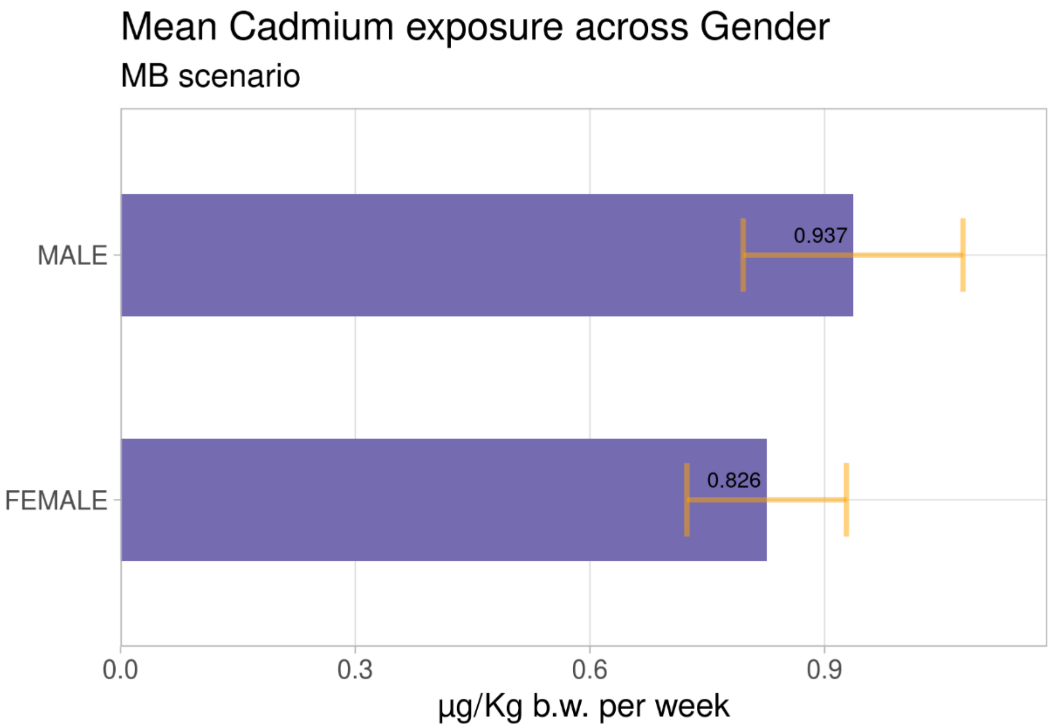

Figure 2. The mean cadmium exposure between the two genders (male and female) at the MB scenario (Kafouris et. 2024).

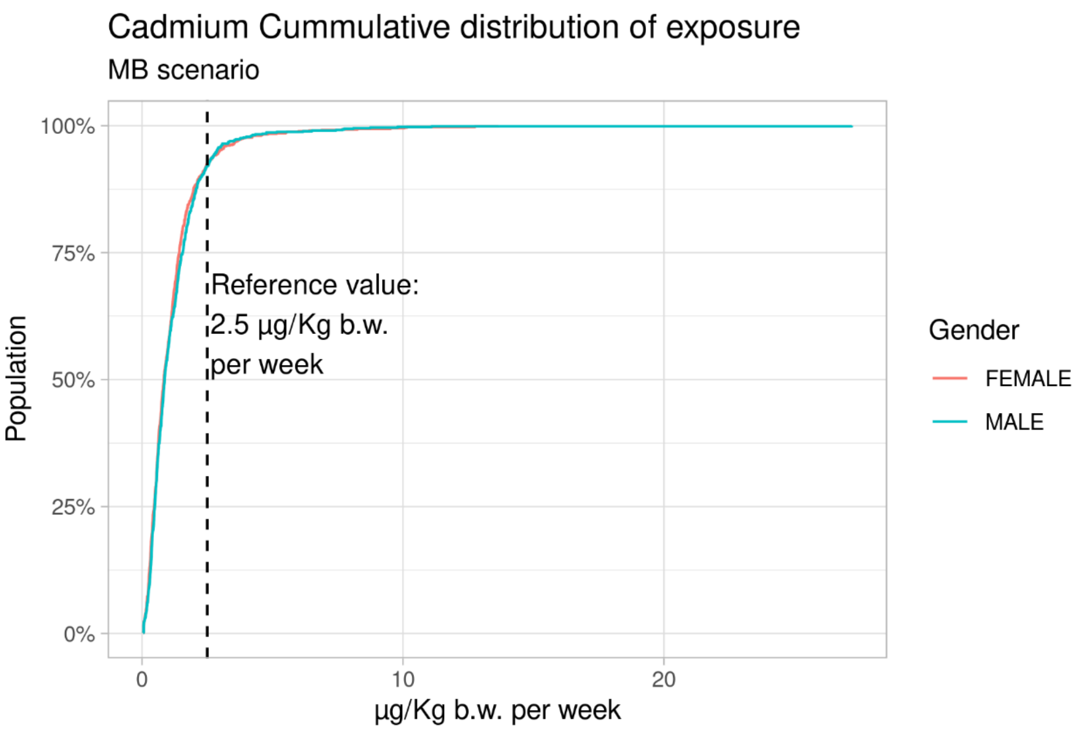

Figure 3. Cumulative distribution of the mean cadmium exposure of the two genders (male and female) at the MB scenario (Kafouris et. 2024).

Table 4. Exposure statistics for cadmium intake (µg/Kg b.w. per week) by Area at the MB scenario.

Area	Min	Max	Mean	95% C.I. *	SD	P25	Median	P75	P95	pctOver**
Ammochostos	0,068	27,196	0,845	(0.541 - 1.148)	1,963	0,298	0,482	0,827	1,936	2.7%
Larnaka	0,053	12,762	0,64	(0.57 - 0.709)	0,674	0,325	0,491	0,745	1,62	1.7%
Lefkosia	0,058	10,156	1,015	(0.813 - 1.217)	1,345	0,356	0,645	1,026	3,13	6.1%
Lemesos	0,063	13,654	0,909	(0.789 - 1.029)	1,055	0,468	0,64	0,915	2,533	6.1%
Pafos	0,046	11,1	0,889	(0.636 - 1.143)	1,161	0,376	0,598	0,859	4,163	6.1%

Note: There is a statistically significant difference in the mean exposure, across the levels of `Area` (ANOVA: F = 3.39, p = 0.01).
* 95% C.I.: 95% Confidence Interval for the mean exposure.
* pctOver: Percentage of population above the reference value.

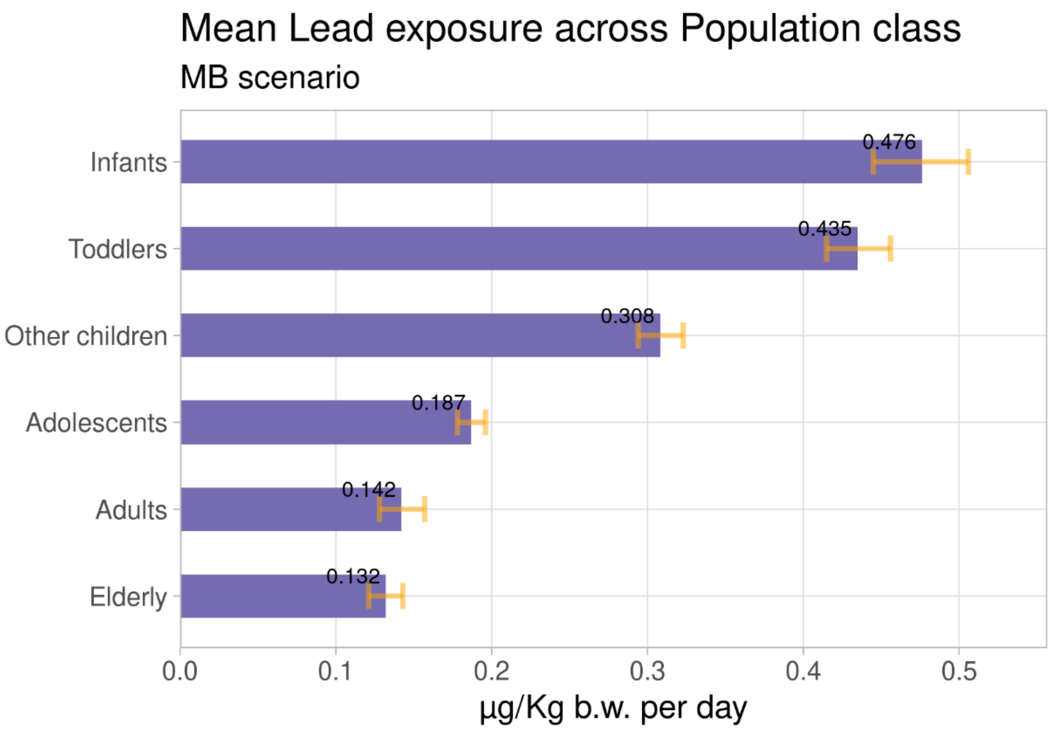

Figure 4. The mean lead exposure between the different population classes at the MB scenario (Kafouris et. 2024).

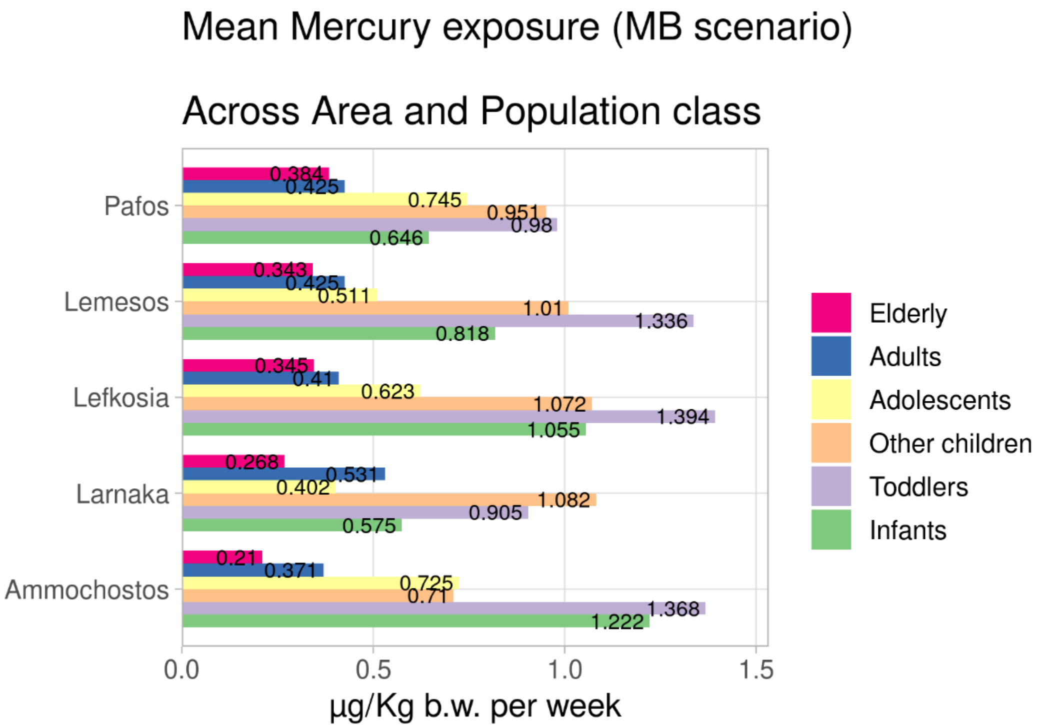

Figure 5. Presentation of the exposure to mercury at the five different areas in Cyprus and across the six population classes (Kafouris et. 2024).

5.3 Exposure at individual level

The estimation of the exposure for each subject in the dietary survey is calculated at the LB, MB and UB scenario and presented in the model. This is derived as a function of a) Consumption Days, b) the total exposure of the subject throughout the consumption days, and c) whether a daily or weekly exposure substance is under study (e.g. Tolerable Weekly Intake means exposure per week, Acceptable Daily Intake mean exposure per day). Users can also “explore” the exposure derived at each food consumption occasion. They can inspect and verify the calculations performed by the model. Specifically, users can identify at which Level of the FoodEx2 hierarchy the exposure was calculated per food consumption occasion and whether the occasion concerned a processed (F28) food.

5.4 Food Contribution to the total exposure

Users can view the food categories contributing to the estimated total exposure of the population. The contribution for each population class in the sample is also calculated. Moreover, users can filter the food items with a minimum percentage (%) of contribution. Any value above 0 is acceptable by the model (Figure 6). ImproRisk also calculates the contribution of food items whose respective codes include a processing facet (F28). Consider, for example, the food item in the occurrence data “Potatoes

{Frying}" with code A00ZT#F28.A07GR, at Level 4 of the FoodEx2 hierarchy. When users choose Level 4, the "Potatoes {Frying}" contribution will be calculated. However, the processing facet is considered only at the relevant Level of the food item (in this case Level 4), as shown in the occurrence data. For example, if the user selects Level 3 (a higher level), the contribution of "Potatoes and similar-" will be calculated that includes the Potatoes {Frying}.

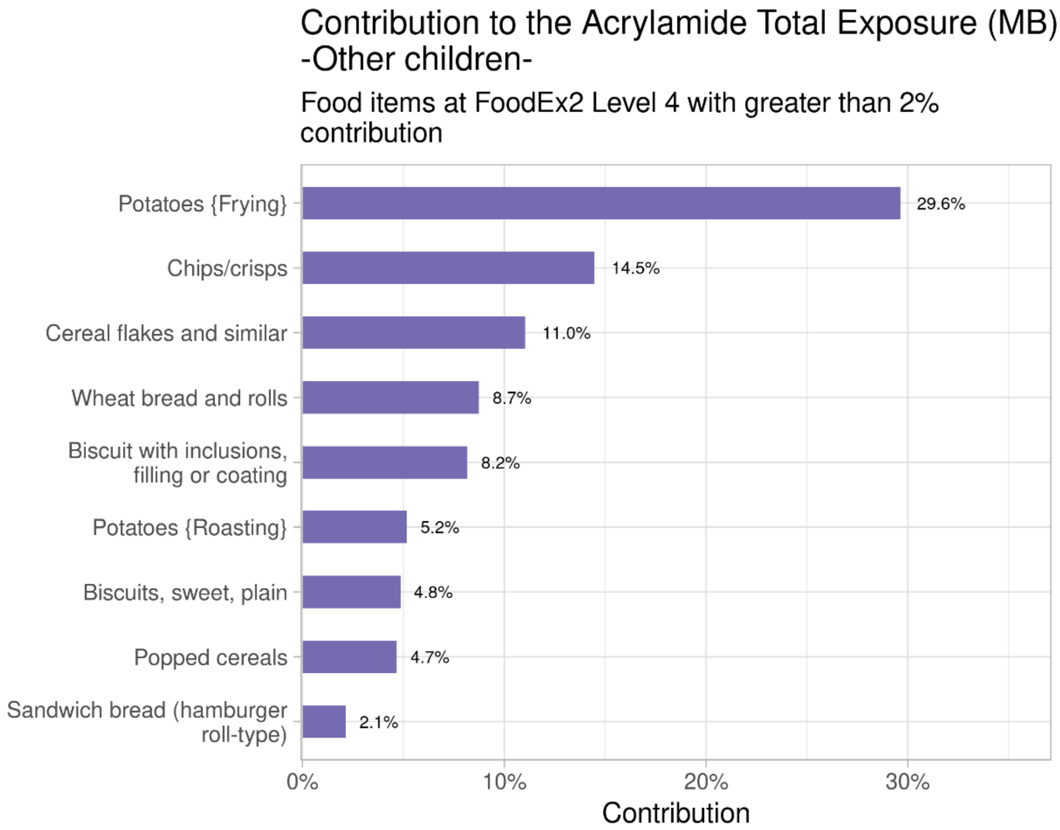

Figure 6. Contribution (>2%) of the food categories to the total acrylamide exposure at the population class of "other children".

6. Reporting

Users of the ImproRisk model can download an automated report in editable Word format. The generated report contains all the basic exposure assessment tables and graphs, the contribution statistics, information on the consumption and occurrence data and several other information regarding the methodology, target population, etc. The report has a pre-fixed structure, yet users can interfere and make any changes in the document that suit their needs.

Acknowledgements

"ImproRisk as an open access Risk Assessment Model" (GP/EFSA/ENREL/2022/03) is a country specific tailor – made project, funded by EFSA under the New Operational Framework between EFSA and the National Focal Points.

Improvast Ltd is also thanked for its contribution to the development of the ImproRisk model.

References

1. European Food Safety Authority, 2015. The food classification and description system FoodEx2 (revision 2). EFSA supporting publication 2015: EN-804, 90pp.

2. European Food Safety Authority, 2014. Guidance on the EU Menu methodology. EFSA Journal 2014;12(12):3944, 80pp.

3. European Food Safety Authority; Management of left-censored data in dietary exposure assessment of chemical substances. EFSA Journal 2010; 8(3), 96pp.

4. Kafouris, D., Christoforou, E., Stephani, D., Sarandi, A., Stavroulakis, G., Christou, E. and Yiannopoulos, S. "Lead, cadmium and mercury determination and human health risk assessment in foods from Cyprus", Journal Food Composition and Analysis, 2024, 128, Article 106007

Abbreviations

Key	Description
BMDL	Benchmark dose lower limit (lower bound of benchmark dose interval)
b.w.	Body weight
LB	Lower bound
LOD	Limit of detection
MB	Middle bound
MoE	Margin of exposure
SD	Standard deviation
P25	25th percentile
P75	75th percentile
P95	95th percentile
UB	Upper bound
wcoeff	Weight coefficient