ImproRisk

v.1.3.1

A model for Dietary Risk Assessment

Description and User Manual

June 2016

Nicosia, Cyprus

ГХК SGL

improvast

Preface

This ImproRisk user manual corresponds with the updated version of ImproRisk (v1.3.1) released in June 2016.

In this new version of the model, some changes and amendments were made, which will be described in detail in the Appendix A. The amendments were prompted by the feedback received from the trainees during the ImproRisk workshop that took place in Larnaca, Cyprus on the 19th May.

Furthermore, some typos within the model and the user manual were corrected and a guide for the COMPREHENSIVE worksheet is added.

Table of Contents

Introduction

The ImproRisk model is a model built in MS Excel, for conducting dietary risk assessment analysis.

It combines Food Consumption dataset with Occurrence (of chemical substances) Dataset and calculates the exposure rates for the population.

It has the ability to support Food Consumption datasets at individual level. Meaning, one can load food consumption data that came out of a food survey. All the individuals are used and the consumption data needs to be coded according to EFSA food categorisation at level 2 (FoodEx level 2) along with the amounts consumed. The food consumption dataset, may contain demographic characteristics of the population and especially **must contain** the individual's own body weight so that the exposure per kilogram of body weight (Kg/ b.w.) is calculated.

Calculation at each food consumption occasion is performed. Every individual, at each food consumption occasion, is exposed to a chemical substance. The exposure at that individual food consumption occasion is calculated by matching the amount of consumption of the food and occurrence of a toxic compound. Individual's body weight is taken into account. Then, by aggregating the results, this leads to the following:

A) Exposure rate for the population. Determine the rate of exposure of the population to a given threshold (i.e. 80% of the population is below the threshold). This can be either weekly threshold, or daily threshold according to the toxic compound and EFSA guidelines.
B) Derivation of (empirical) probability and cumulative distributions of the exposures. This will enable the risk manager to have a clearer view on how the exposure is distributed over the population of interest.
C) Explore the effect of demographic characteristics on the exposure. For example, one can compare the exposure towards gender or age group of the population.

The above can be derived for all Lower Bound (LB), Middle Bound (MB), and Upper Bound (UB) scenarios of the occurrence dataset and the exposure is calculated at the FoodEx level 2.

The occurrence dataset needs to be in FoodEx Level 2, and the Consumption dataset must be in FoodEx level 4. Consumption at higher levels (i.e. Level 3, Level 2) can also be supported with minor modifications.

Benefits

I) The calculation is performed at the individual level, thus enabling the estimation of exposure distribution and waving the biased effect of the mean consumption.

II) It is not a closed box model. All calculations are there to inspect. The formulas and all the methodology are transparent so the model results can be validated easily.

III) It is a straightforward and user friendly model.

IV) Free to use.

Ownership and intellectual rights

ImproRisk is owned by the State General Laboratory of Cyprus (SGL) and was developed by the private company Improvast. SGL reserves the property rights. The model is free to use, but any unauthorised modification is prohibited under the Cyprus law.

Contact

This manual was prepared by Improvast. Should you find any errors or have any comments/suggestions please contact info@improvast.com.

Additionally, should you have any suggestions for further development or encounter bugs please report them to gstavroulakis@sgl.moh.gov.cy or info@improvast.com.

We welcome all feedback from all interested parties.

Methodology

The model takes into account the mean occurrence (LB, MB, & UB scenario) and uses the consumption data at an individual level.

The calculated individual exposure, can be later be used to cross tabulate by Gender, Age group and Area.

The model is ready to calculate the intake levels of any substance under study by substituting the occurrence data via an automated VBA code (Visual Basic for Applications). Moreover, the model enables the user to easily install a new Consumption Database.

<u>The Exposure Daily Intake of Toxic Substance(EDI) is calculate as follows:</u>

a) For every single food consumption occasion, at any day, of any food category of food consumption the following formula is applied:

$$\frac{Consumption\ (grams\ of\ food)\text{x}Occurence\left({}^{mg}/_{Kg\ of\ food}\right)}{Body\ Weight\ (Kg)} = Exposure\ ({}^{\mu g}/_{Kg\ b.w.})$$

b.w. = body weight

b) The body weight is each individual's body weight. The final outcome of the equation is the exposure (in µg/Kg b.w.) for each instance the subject consumes each and every specific food.

c) The Exposure:

The Exposure for each instance calculated in step a) is summed for all days and for every food category in FoodxLevel2 (160 categories).

To calculate the expected daily intake (EDI), we take the average throughout the number of days of the study.

This results in the EDI in µg/Kg b.w. for each and every one of the individuals in the food survey

Note that 3 types of EDI's are calculated – LB, MB, & UB.

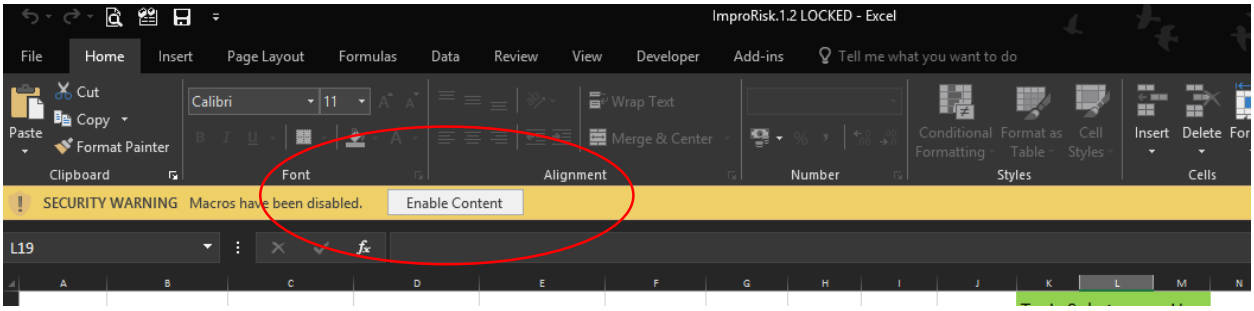
This way the exposure rate is established for each individual at an Optimistic (LB), an "Average" (MB) and a Pessimistic (UB) scenario.

Launching the ImproRisk

Open the file ***ImproRisk.1.3.xlsx***

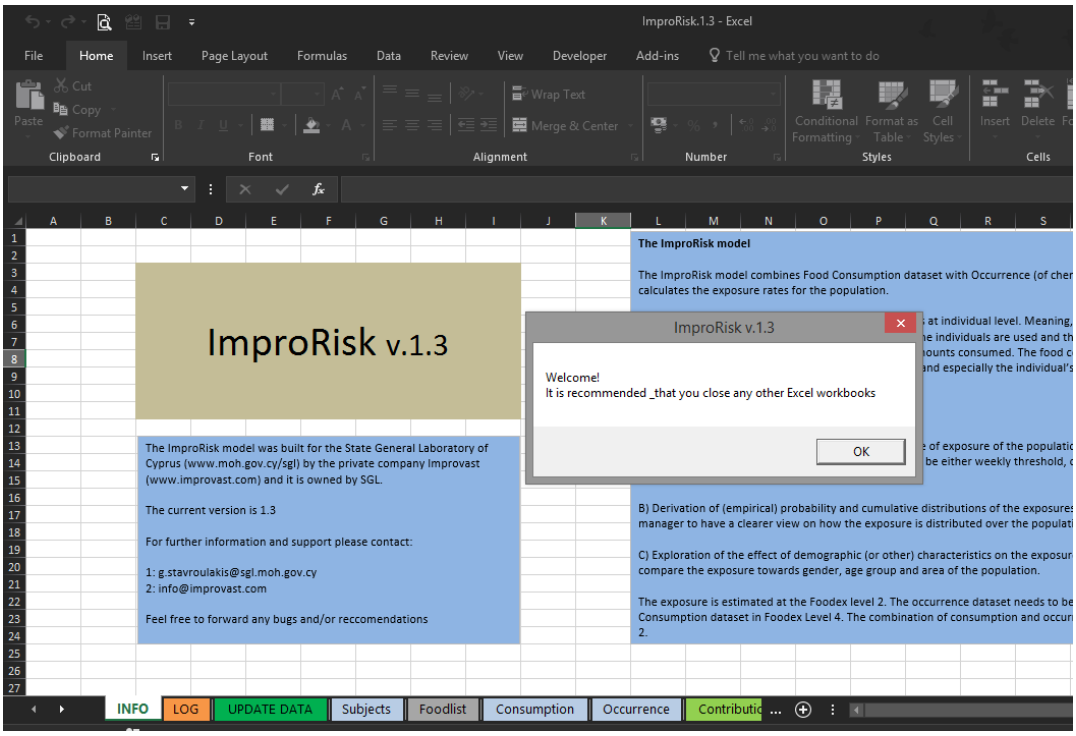
If you see the security warning (Figure 1) that macros have been disabled, then click on the Enable Content button. Macros are programming code in VBA which is included in Excel. The Macros must be activated for the model to function.

Figure 1: Security warning for enabling macros in Excel



The user will get a welcoming message informing you to close all other Excel workbooks. This will reduce the processing speed and increase the efficiency of the model.

Figure 2: Opening ImproRisk 1.3



WORKSHEETS of the ImproRisk model

An overview of each worksheet's functionality in the model reporting.

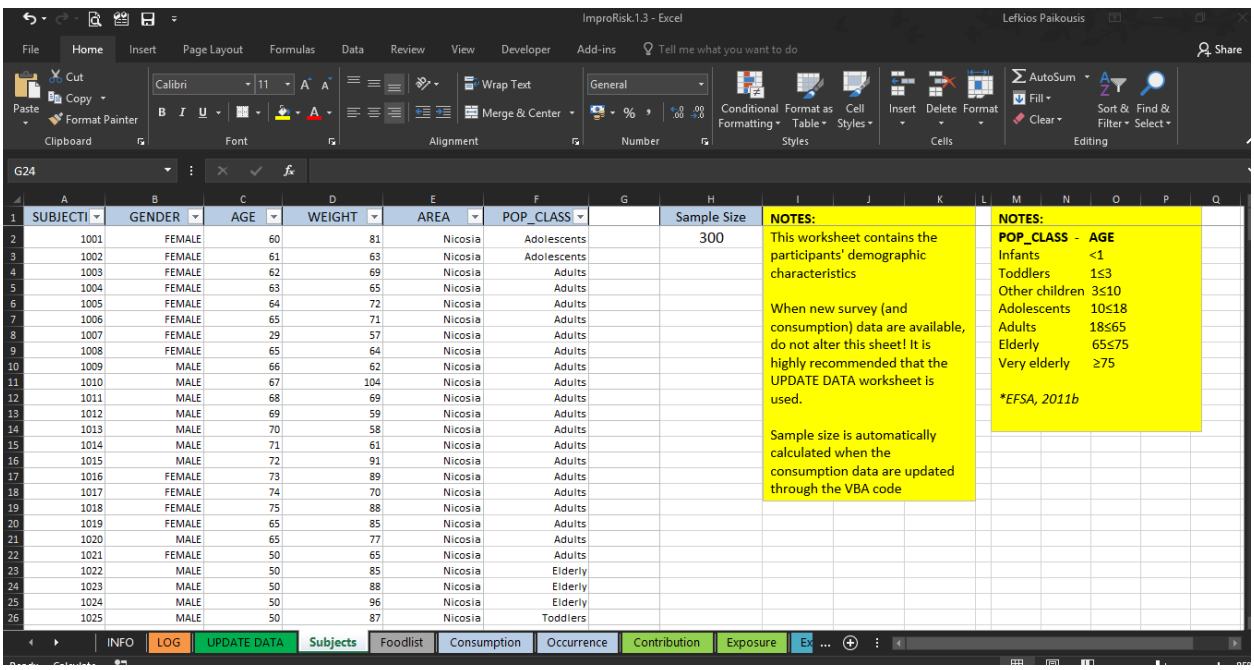
Worksheet **Subjects**

Press CTRL-Home to navigate to the TOP LEFT cell (A1)

The worksheet Subjects tabulates the sample population data to be used. Each line represents an individual. Across the worksheet the user can see the following: The subject ID, Gender, Age, Weight, Area and Population Class (POP_CLASS).

The *Sample Size* is calculated automatically via the VBA code when the *Consumption and Subjects* dataset is installed (see *VBA code*)

Figure 3: Screenshot of the Subjects worksheet



Worksheet **Foodlist**

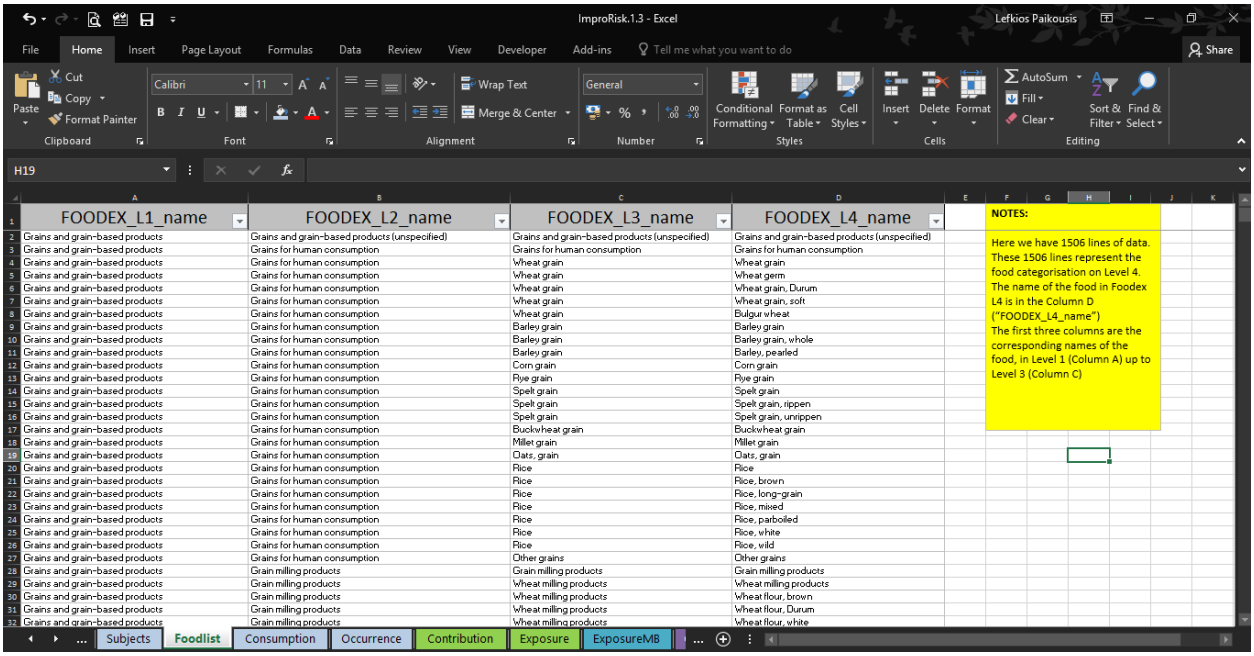
Press CTRL-Home to navigate to the TOP LEFT cell (A1)

This worksheet contains the food categorisation from FoodEx level 1 to FoodEx Level 4 as indicated by EFSA. The name of the food in FoodEx L4 is in the Column D ("FOODEX_L4_name"). The first three columns are the corresponding names of the food, in Level 1 (Column A) up to Level 3 (Column C).

Press CTRL-Down Key to navigate to the END of the worksheet.

Press CTRL-Home to navigate back to the beginning

Figure 4: Screenshot of the worksheet Foodlist



Worksheet **Consumption**

Press CTRL-Home to navigate to the TOP LEFT cell (A1)

The first five columns (A, B, C, D, E, coloured in light BLUE background colour) contain the uploaded consumption dataset.

Press CTRL-Down Key

Thousands of observations can be observed in this worksheet. These are the **food consumption occasions** within the sample, i.e. the *SUBJECTID* of the person who consumed the food, the sequential *DAY* of consumption in the food survey (what day the individual consumed it), the *amount of food* (in grams) and the *FoodEx level 4 name* of the food. Subjects consumed different and various amounts and type of foods at each day, so the user will see the same *SUBJECTID* and *DAY* many times.

Press CTRL-Home to navigate back to the beginning

The next two columns "F" and G" (*grey coloured*) translate the FoodEx L4 name into FoodEx L1 and L2 respectively. This is what the worksheet **Foodlist** is used for.

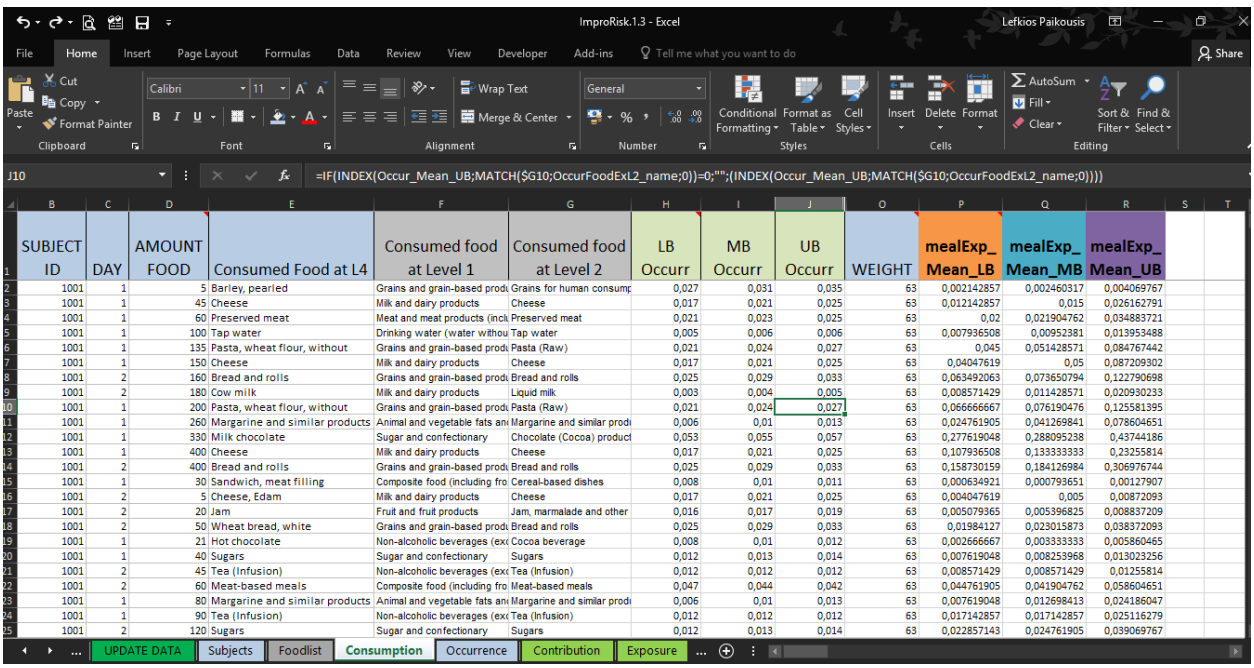
The next 3 columns "H", "I" and J" hold the occurrence of the toxic in the LB, MB, and UB scenario respectively. These values are obtained (automatically) from the

Worksheet Occurrence that we will be presented later in this user manual.

The next 5 columns “K”, “L”, “M”, “N”, “O” hold the demographic characteristics of the subject of that particular food consumption occasion; Gender, Area, Pop Class, Age and Weight.

The WEIGHT will be used in the Exposure calculation. The other demographics are not needed for the exposure but this worksheet serves as a standalone dataset for further analysis either within the ImproRisk or in an outside data processing software i.e. SPSS, R etc.

Figure 5: Screenshot of worksheet Consumption



SUBJECT ID	DAY	AMOUNT FOOD	Consumed Food at L4	Consumed food at Level 1	Consumed food at Level 2	LB Occurr	MB Occurr	UB Occurr	WEIGHT	mealExp_ Mean_LB	mealExp_ Mean_MB	mealExp_ Mean_UB
1001	1	5	Barley, pearled	Grains and grain-based produ	Grains for human consump	0,027	0,031	0,035	63	0,002142857	0,002460317	0,004069767
1001	1	45	Cheese	Milk and dairy products	Cheese	0,017	0,021	0,025	63	0,012142857	0,015	0,026162791
1001	1	60	Preserved meat	Meat and meat products (incl	Preserved meat	0,021	0,023	0,025	63	0,02	0,021904762	0,034883721
1001	1	100	Tap water	Drinking water (water withou	Tap water	0,005	0,006	0,006	63	0,007936508	0,00952381	0,013953488
1001	1	135	Pasta, wheat flour, without	Grains and grain-based produ	Pasta (Raw)	0,021	0,024	0,027	63	0,045	0,051428571	0,084767442
1001	1	150	Cheese	Milk and dairy products	Cheese	0,017	0,021	0,025	63	0,04047619	0,05	0,087209302
1001	2	160	Bread and rolls	Grains and grain-based produ	Bread and rolls	0,025	0,029	0,033	63	0,063492063	0,073650794	0,122790698
1001	2	180	Cow milk	Milk and dairy products	Liquid milk	0,003	0,004	0,005	63	0,008571429	0,011428571	0,020930233
1001	1	200	Pasta, wheat flour, without	Grains and grain-based produ	Pasta (Raw)	0,021	0,024	0,027	63	0,066666667	0,076190476	0,125581395
1001	1	260	Margarine and similar products	Animal and vegetable fats an	Margarine and similar produ	0,006	0,01	0,013	63	0,024761905	0,041269841	0,078604651
1001	1	330	Milk chocolate	Sugar and confectionary	Chocolate (Cocoa) product	0,053	0,055	0,057	63	0,277619048	0,288095238	0,43744186
1001	1	400	Cheese	Milk and dairy products	Cheese	0,017	0,021	0,025	63	0,107936508	0,133333333	0,23255814
1001	2	400	Bread and rolls	Grains and grain-based produ	Bread and rolls	0,025	0,029	0,033	63	0,158730159	0,184126984	0,306976744
1001	1	30	Sandwich, meat filling	Composite food (including fro	Cereal-based dishes	0,008	0,01	0,011	63	0,000634921	0,000793651	0,00127907
1001	2	5	Cheese, Edam	Milk and dairy products	Cheese	0,017	0,021	0,025	63	0,004047619	0,005	0,00872093
1001	2	20	Jam	Fruit and fruit products	Jam, marmalade and other	0,016	0,017	0,019	63	0,005079365	0,005396825	0,008837209
1001	2	50	Wheat bread, white	Grains and grain-based produ	Bread and rolls	0,025	0,029	0,033	63	0,01984127	0,023015873	0,038372093
1001	1	21	Hot chocolate	Non-alcoholic beverages (exc	Cocoa beverage	0,008	0,01	0,012	63	0,002666667	0,003333333	0,005860465
1001	1	40	Sugars	Sugar and confectionary	Sugars	0,012	0,013	0,014	63	0,007619048	0,008253968	0,013023256
1001	2	45	Tea (Infusion)	Non-alcoholic beverages (exc	Tea (Infusion)	0,012	0,012	0,012	63	0,008571429	0,008571429	0,01255814
1001	2	60	Meat-based meals	Composite food (including fro	Meat-based meals	0,047	0,044	0,042	63	0,044761905	0,041904762	0,058604651
1001	1	80	Margarine and similar products	Animal and vegetable fats an	Margarine and similar produ	0,006	0,01	0,013	63	0,007619048	0,012698413	0,024186047
1001	1	90	Tea (Infusion)	Non-alcoholic beverages (exc	Tea (Infusion)	0,012	0,012	0,012	63	0,017142857	0,017142857	0,025116279
1001	2	120	Sugars	Sugar and confectionary	Sugars	0,012	0,013	0,014	63	0,022857143	0,024761905	0,039069767

Go to columns “P”, “Q” and “R”

This is the calculation of *exposure* at each food consumption occasion for the **LB**, MB and **UB** occurrence scenario respectively.

The *exposure* is calculated as follows:

$$\text{Exposure } (\mu g/\text{kg b.w} *) = \text{Consumption (grams)x Occurrence (mg/Kg of food)} \Big/ \text{Weight(kg)}$$

*b.w.= Body Weight

The outcome is exposure in *μg* per *Kg of body weight* of the individual

Figure 6: Calculation of exposure at each food consumption occasion

SUM =IFERROR($D8*I8/O8;"")

	A	B	C	D	E	F	G	H	I	J	O	P	Q	R
1	AA	SUBJECT ID	DAY	AMOUNT FOOD	Consumed Food at L4	Consumed food at Level 1	Consumed food at Level 2	LB Occurr	MB Occurr	UB Occurr	WEIGHT	mealExp_ Mean_LB	mealExp_ Mean_MB	mealExp_ Mean_UB
2	10	1001	1	5	Barley, pearled	Grains and grain-based produ	Grains for human consump	0,027	0,031	0,035	63	0,002142857	0,002460317	0,004069767
3	14	1001	1	45	Cheese	Milk and dairy products	Cheese	0,017	0,021	0,025	63	0,012142857	0,015	0,026162791
4	2	1001	1	60	Preserved meat	Meat and meat products (incl	Preserved meat	0,021	0,023	0,025	63	0,02	0,021904762	0,034883721
5	7	1001	1	100	Tap water	Drinking water (water withou	Tap water	0,005	0,006	0,006	63	0,007936508	0,00952381	0,013953488
6	4	1001	1	135	Pasta, wheat flour, without	Grains and grain-based produ	Pasta (Raw)	0,021	0,024	0,027	63	0,045	0,051428571	0,084767442
7	5	1001	1	150	Cheese	Milk and dairy products	Cheese	0,017	0,021	0,025	63	0,04047619	0,05	0,087209302
8	19	1001	2	160	Bread and rolls	Grains and grain-based produ	Bread and rolls	0,025	0,029	0,033	63	0,063492063	=IFERROR($D8*I8/O8;"")	
9	25	1001	2	180	Cow milk	Milk and dairy products	Liquid milk	0,003	0,004	0,005	63	0,008571429	IFERROR(value; value_if_error)	
10	9	1001	1	200	Pasta, wheat flour, without	Grains and grain-based produ	Pasta (Raw)	0,021	0,024	0,027	63	0,066666667	0,076190476	0,125581395
11	1	1001	1	260	Margarine and similar products	Animal and vegetable fats an	Margarine and similar produ	0,006	0,01	0,013	63	0,024761905	0,041269841	0,078604651
12	3	1001	1	330	Milk chocolate	Sugar and confectionary	Chocolate (Cocoa) product	0,053	0,055	0,057	63	0,277619048	0,288095238	0,43744186
13	6	1001	1	400	Cheese	Milk and dairy products	Cheese	0,017	0,021	0,025	63	0,107936508	0,133333333	0,23255814
14	24	1001	2	400	Bread and rolls	Grains and grain-based produ	Bread and rolls	0,025	0,029	0,033	63	0,158730159	0,184126984	0,306976744

IMPORTANT NOTE 1:

1. Calculation of exposure at each food consumption occasion:

The excel built in formula IFERROR() is used when the multiplication is performed.

In case where an invalid Food name is used for a food consumption occasion, then no occurrence value for this consumption occasion will be found, thus an error is created in the columns of Exposure (P, Q, and R). This way, if an invalid name is used, then the model assumes that there is no food consumption occasion (leaving the exposure blank) thus no exposure for the particular food consumption occasion, and the aggregating statistics calculations (sheet Exposure) are performed without any interruption.

2. The occurrence of each food consumption occasion is retrieved in columns H, I, J. When the value is empty, this means that no mean occurrence for that food category is determined in the "**Occurrence**" sheet.

Solution:

By using the filters provided by Excel, any errors can be identified easily e.g. consumption name not included at the food level 2 or occurrence data values do not exist in columns H,J and I.

IMPORTANT NOTE 2:

The consumption database, when in need to be updated, it must be done with a template that is provided (Subjects_Consumption_Template.xlsx). When the template is used, then use the UPDATE DATA worksheet, where VBA code lies for updating Consumption data. The VBA code will ask the user to locate and select the file in the local disk and replacement of both the Subject information and Consumption information will be performed automatically.

Note that the template name is not of importance. Name the resulting file appropriately for easier reference. What is of importance, it is the name of the two worksheets within the template that need to be named “Subjects” for the subject information and “Consumption” for the Consumption information

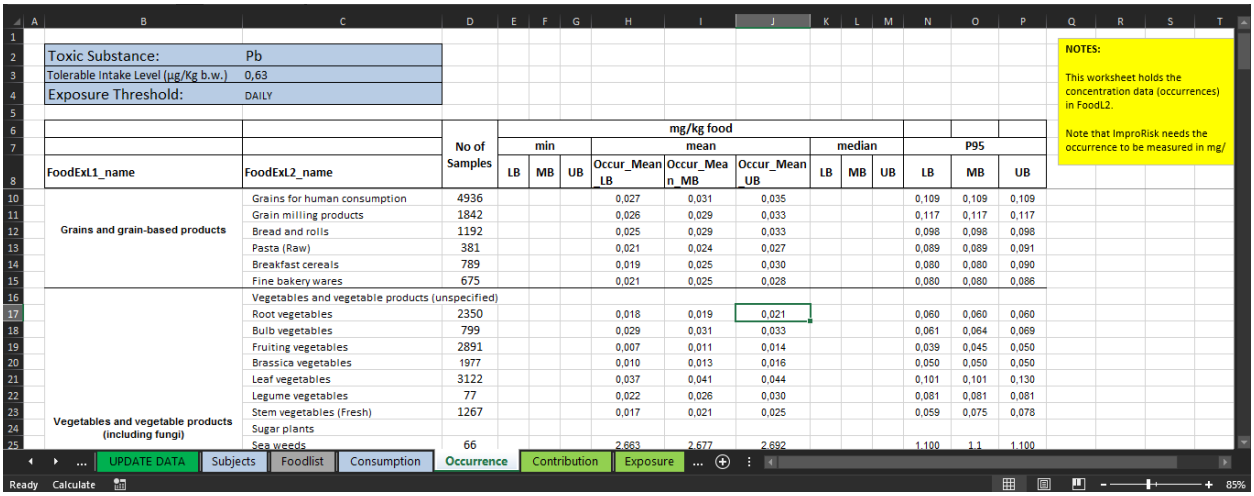
See instructions in the Chapter

Worksheet **Occurrence**

This worksheet contains the concentration data (occurrences) in FoodL2.

Note that ImproRisk needs the occurrence to be measured in mg/ Kg of food

Figure 7: Screenshot for the worksheet Occurrence



Toxic Substance:	Pb
Tolerable Intake Level (µg/Kg b.w.)	0,63
Exposure Threshold:	DAILY

NOTES:

This worksheet holds the concentration data (occurrences) in FoodL2.

Note that ImproRisk needs the occurrence to be measured in mg/

			mg/kg food											
			min			mean			median			P95		
FoodExL1_name	FoodExL2_name	No of Samples	LB	MB	UB	Occur_Mean_LB	Occur_Mean_MB	Occur_Mean_UB	LB	MB	UB	LB	MB	UB
Grains and grain-based products	Grains for human consumption	4936				0,027	0,031	0,035				0,109	0,109	0,109
	Grain milling products	1842				0,026	0,029	0,033				0,117	0,117	0,117
	Bread and rolls	1192				0,025	0,029	0,033				0,098	0,098	0,098
	Pasta (Raw)	381				0,021	0,024	0,027				0,089	0,089	0,091
	Breakfast cereals	789				0,019	0,025	0,030				0,080	0,080	0,090
	Fine bakery wares	675				0,021	0,025	0,028				0,080	0,080	0,086
Vegetables and vegetable products (including fungi)	Vegetables and vegetable products (unspecified)													
	Root vegetables	2350				0,018	0,019	0,021				0,060	0,060	0,060
	Bulb vegetables	799				0,029	0,031	0,033				0,061	0,064	0,069
	Fruiting vegetables	2891				0,007	0,011	0,014				0,039	0,045	0,050
	Brassica vegetables	1977				0,010	0,013	0,016				0,050	0,050	0,050
	Leaf vegetables	3122				0,037	0,041	0,044				0,101	0,101	0,130
	Legume vegetables	77				0,022	0,026	0,030				0,081	0,081	0,081
	Stem vegetables (Fresh)	1267				0,017	0,021	0,025				0,059	0,075	0,078
	Sugar plants													
	Sea weeds	66				2,663	2,677	2,692				1,100	1,1	1,100

IMPORTANT NOTE:

In case, the values have to be updated, it must be done with a template that is provided (Occurrence_Template.xlsx). Whenever the template is used, it is essential to use the UPDATE DATA worksheet, where VBA code lies for updating occurrence data. The VBA code will ask the user to locate and select the file in the local disk and the replacement of the values will be performed automatically.

Note that the template name is not of any importance. Name the resulting workbook appropriately for easier reference. What is of importance, it is the name of the worksheet that needs to "Sheet1".

See instructions in the Chapter

Updating the Occurrence and

Worksheet **Contribution**

The worksheet Contribution presents the calculated exposure aggregated by each Food level 1 and Food level 2 category. The aggregation is the summation of the exposure for each food consumption occasion.

Eventually, the percent contribution (%) (for LB, MB and UB scenario) of each food category to the total exposure of the entire sample.

Tables in the worksheet

Table 1: Total Exposure (µg/Kg of body weight) over the total population broken down by FoodEx L1 name.

Table 2: Contribution of each food category (Level 1) on the Total Exposure (µg/Kg of body weight) over the total population.

Table 3: Mean exposure per Day per Person by each food category (Level 1).

This is the values of Table 1 divided by the *Number of participants* and then by the number of food survey days (*NDays*)

Table 4: Total Exposure (µg/Kg of body weight) over the total population broken down by FoodEx L2 name and Contribution of each food category (Level 2) on the Total Exposure (µg/Kg of body weight) over the total population.

Tables can then be **sorted** in the following manner (Figure 8): e.g. for Table 1:

1. Select the whole table including the titles (FoodEx Level 1, LB, MB, UB)

2. On the excel menu bar ribbon (horizontal menu) go to **Data** and then click on the **Sort** button

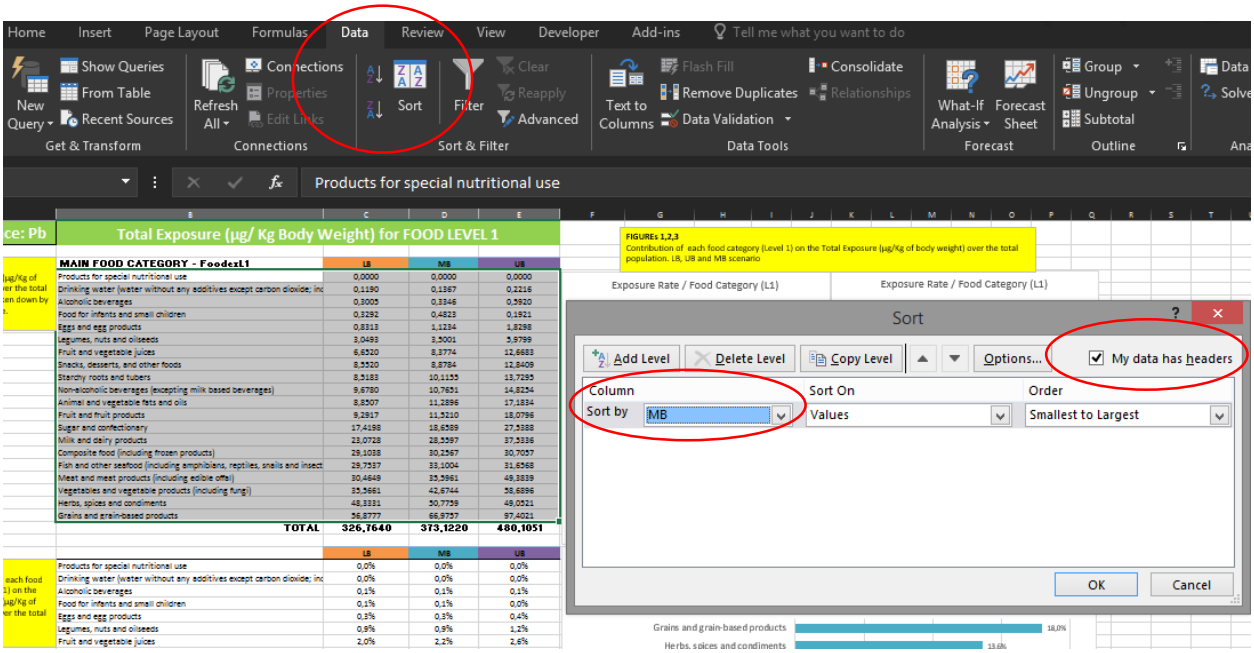
3. Make sure the "My data has headers" is enabled (clicked)

4. Select the **Sort by** title [Food name or just the LB, MB or UB values]

5. Press OK

6. Press *Shift+F9* to recalculate the worksheet

Figure 8: Sort the contribution by Food name or Scenario value



Figures in the worksheet

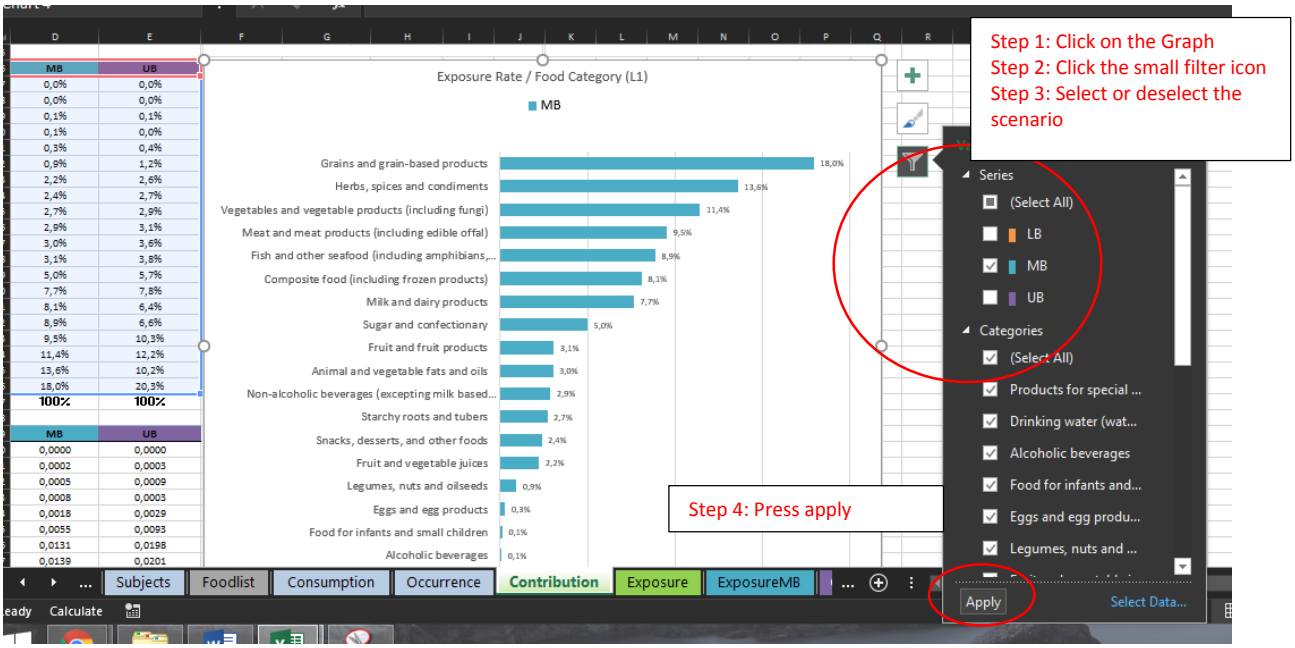
Figures 1, 2,3: Visualisation of the contribution for the Level 1 categories for the LB, MB and UB scenarios.

Figure 4: Visualisation of the contribution for the Level 2 categories

Figure 4, shows the contribution of each food FoodEx L2 category to the Total Exposure (µg/Kg of body weight) over the total population, at the MB scenario.

Users with Excel 2013 or later versions can filter the visualisation by LB and UB scenarios as well (Figure 9).

Figure 9: How to view the food contribution in LB, MB and UB scenario



Worksheet **Exposure**

In this worksheet, the exposure assessment (µg/Kg body weight) is performed for each individual (subject) in the food survey.

The total exposure for each subject that participated in the food study, is calculated for:

a. The total survey duration (NDay)

b. Daily or Weekly depending on the tolerable threshold type of the toxic

The DAILY or WEEKLY exposure is calculated using the following formula (see Figure 10):

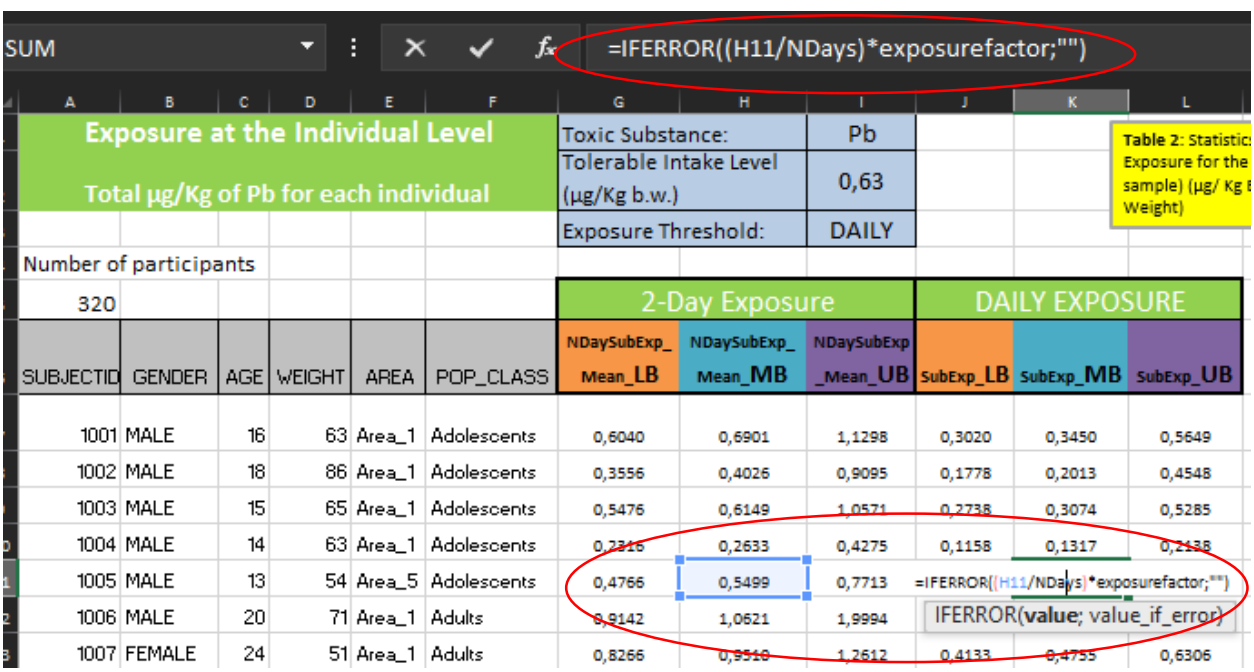
(NDay-Exposure/ NDays) x Exposurefactor

where:

exposurefactor is either 1 if DAILY threshold type, or 7 if WEEKLY threshold type. This value is set via the VBA code during updating of the Occurrence data

NDays = Number of survey Days (captured by the VBA code when updating the Consumption data)

Figure 10: Calculation of DAILY or WEEKLY exposure of an individual



Eventually, if the threshold type is DAILY, then DAILY Exposure = (NDay-Exposure/ NDays)*1. If the threshold type is WEEKLY, then the WEEKLY EXPOSURE = (NDay-Exposure/NDays)*7

The demographics of the subjects are obtained (via excel formulas) from the worksheet **Subjects**. This way, the whole table can serve as a separate dataset for processing in an another software e.g. SPSS, R etc.

Worksheet **MeanConsumption**

In this worksheet, calculations are performed in order the estimate the average chronic consumption (Figure 11).

TABLES in the worksheet:

TABLE 1: First, the average consumption (throughout the survey period) **for each individual** and for each food category in Level 1, is calculated. The *SubjectID* (individuals) is on the vertical axis and the 20 food categories (Level 1) on the horizontal axis.

The average intake of each individual is taken over the N day study period. The average value for each subject is then considered when calculating the "chronic" average consumption[1].

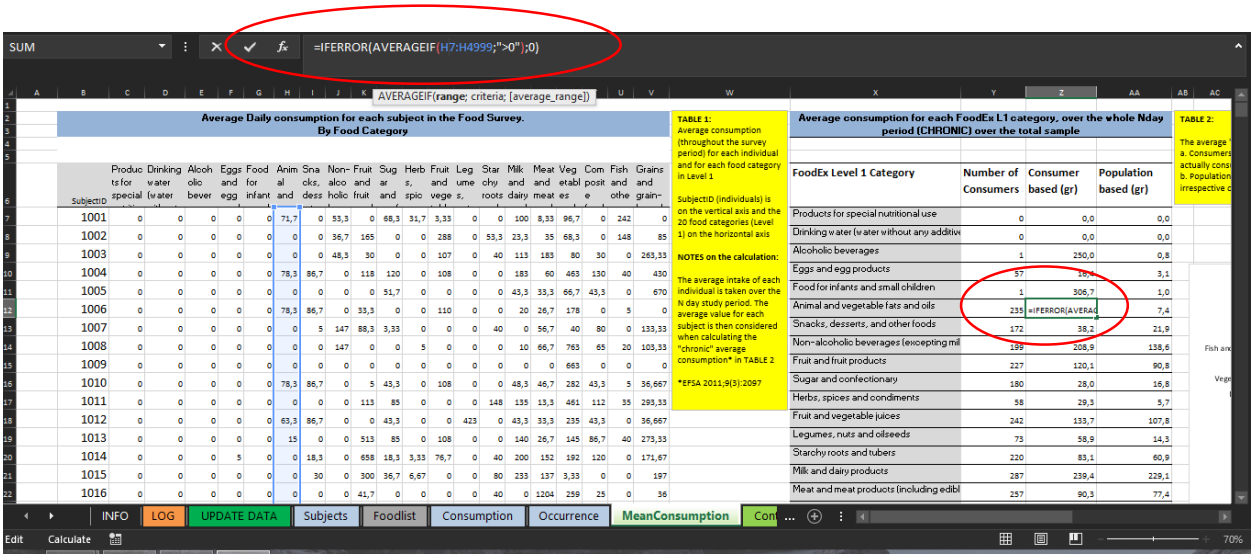
TABLE 2: The average "chronic" consumption for:

a. Consumers only, is calculated by taking the average of those who actually consumed the food.

b. Population based, is calculated by taking the average of all individuals irrespective of their zero consumption (see Figure 11 below).

[1] EFSA 2011;9(3):2097

Figure 11: Calculation of chronic average consumption. Consumer based and Population based



Worksheet **ExposureMB**

In this worksheet, a probability (histogram) and a cumulative distribution of the MB exposure is created.

To create the histogram, the table of frequencies is created first.

Figure 12: Screenshot of worksheet ExposureMB

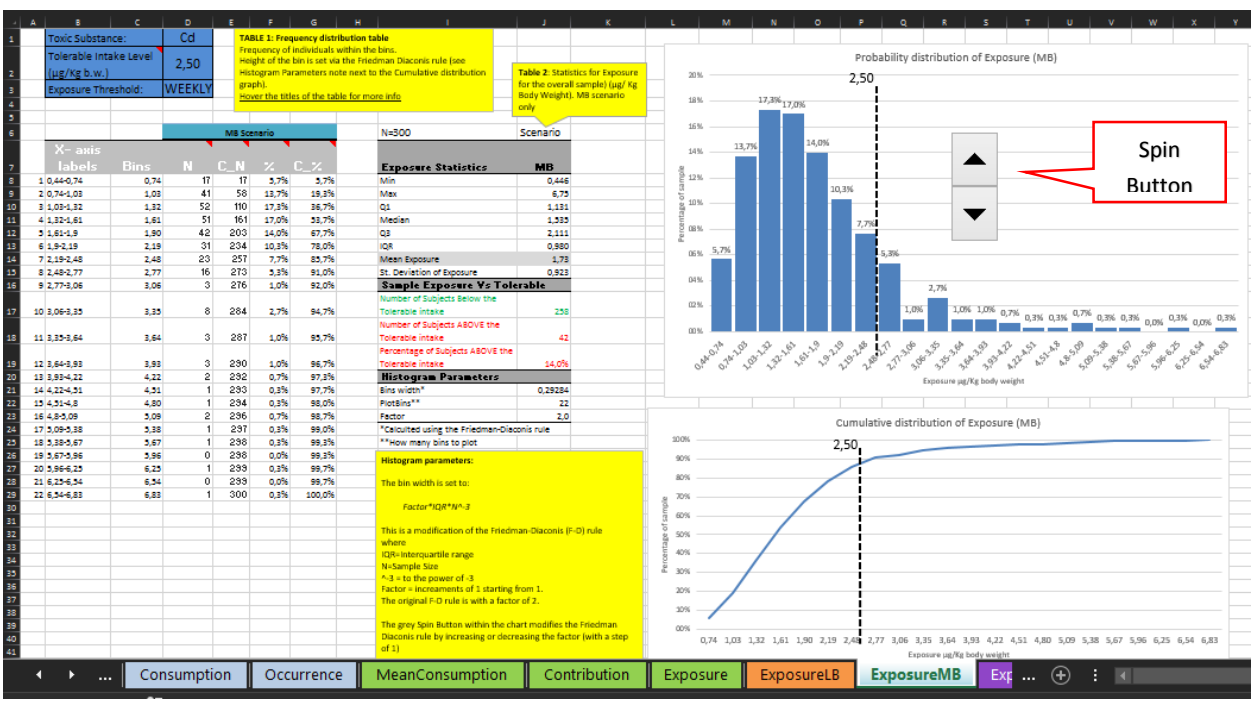


TABLE 1: Frequency distribution table

Frequency of individuals within the bins.

Width of the bin is set via the Friedman-Diaconis (F-D)[2] rule. The F-D rule works well in practice (Scott, D. (1992))

The bin width is set to:

$$Factor*IQR*N^{-3}$$

This is a modification of the Friedman-Diaconis rule

where

IQR=Interquartile range

N=Sample Size

Factor = increments of 1 starting from 1.

The original F-D rule is with a factor of 2.

Using the grey Spin Button (see Figure 12) within the chart enables the user to modify the Friedman-Diaconis rule by increasing or decreasing the factor (with a step of 1). This will change the height of the bin and get a better visualisation of the histogram.

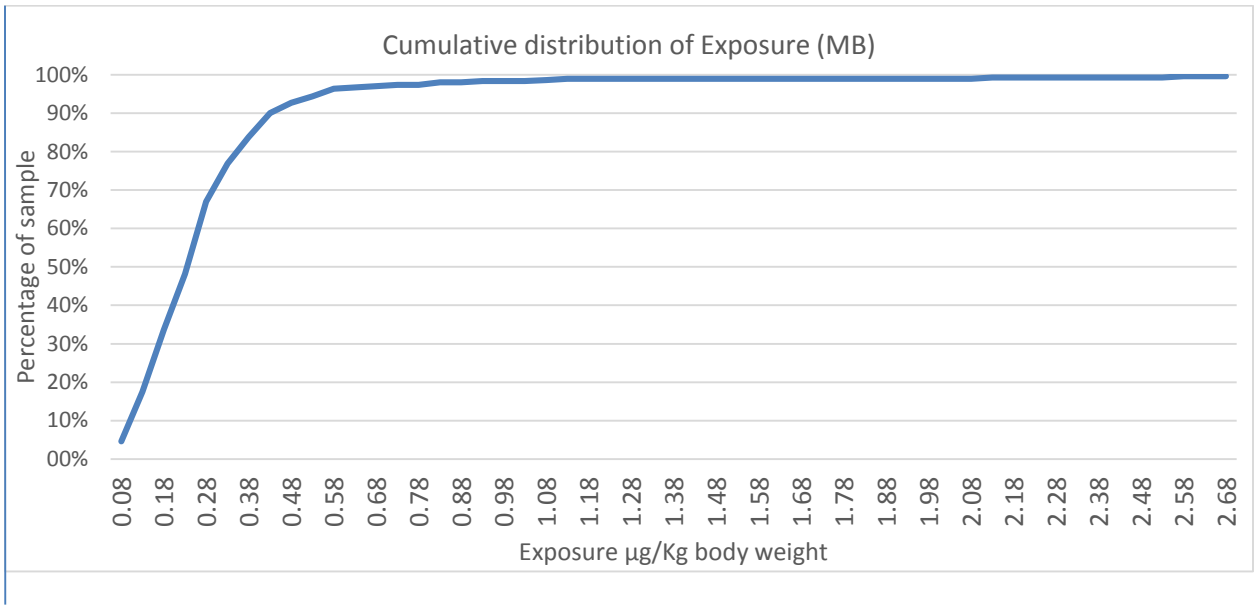
Then, the number of Bins is set to:

$$(Max-Min)/binswidth$$

The Max is the Maximum observed value of the exposures across the population.

[2] https://en.wikipedia.org/wiki/Freedman%E2%80%93Diaconis_rule

Figure 13: Example of a cumulative distribution of exposure



Worksheet **ExposureLB**

The exact same calculations and graph reporting as in ExposureMB but for the LB scenario of exposure.

Worksheet **ExposureUB**

The exact same calculations and graph reporting as in ExposureMB but for the UB scenario of exposure.

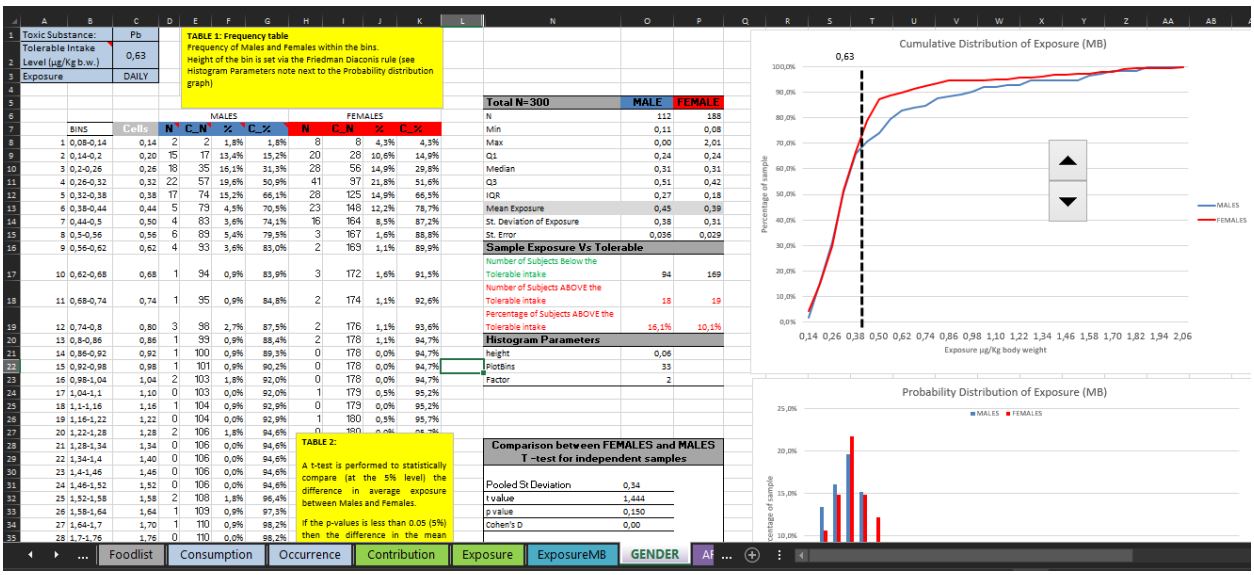
Worksheet **GENDER**

The bin width is the same as the width already *set in the Exposure MB*.

The statistics needed for the calculation of the height are taken those of the MB scenario i.e. IQR, Sample SIze, Max and Min exposure.

The grey Spin Button within the chart modifies the factor of the modified Friedman Diaconis rule by increasing or decreasing it (with a step of 1) as with the ExposureMB worksheet.

Figure 14: Screenshot of the worksheet GENDER



Comparison between FEMALES and MALES

A t-test (Figure 15) is performed to statistically compare (at the 5% level) the difference in average exposure between Males and Females.

If the p-values is less than 0.05 (5%) then the difference in the mean exposure between Males and Females of the sample can be generalised to the population under study (i.e. statistically significant).

Pooled Standard deviation is calculated using the following formula

$$Sp = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

where *subscript 1 for Male, subscript 2 for Female, S= Standard Deviation and n=sample size*

t-value[3] is calculate using the following formula

$$t = \frac{Mean\ Exposure\ Males - Mean\ Exposure\ Females}{Sp * \sqrt{({}^{1}/_{Nmales} + {}^{1}/_{Nfemales})}}$$

and p value[4] is calculated using built in functions in Excel and using two tailed test.

Result of the t-test is reported (Figure 15).

Figure 15: T-test for the comparison of the mean exposure (MN scenario) between males and females

Comparison between FEMALES and MALES T -test for independent samples	
Pooled St Deviation	0,74
t value	2,089
p value	0,037
Cohen's D	0,05
Result There is a statistically significant difference in the average µg/Kg b.w. Intake between Males and Females	

COHEN's D

Cohen's D (Cohen, J. (1977) is a measure of quantification of the difference (effect size) between the Males and Females.

Cohen's D guidelines:

D ~ 0,20 -> small difference

D ~ 0,50 -> moderate difference

D > 0,80 -> large difference

For example, if D is more than 0.80 then the observed difference is considered to be large.

The model produces comparative graphs for Male and Female such as probability and cumulative distribution of exposure (Figure 16) as well as a comparison graph of the proportion of the sample that exceed the tolerable threshold (Figure 17).

[3] https://onlinecourses.science.psu.edu/stat200/node/60

[4] pvalue=2*T. DIST.RT(ABS(t_value) ;(N_MALES+N_FEMALES-2))

Figure 16: Cumulative distribution of the exposure by Gender

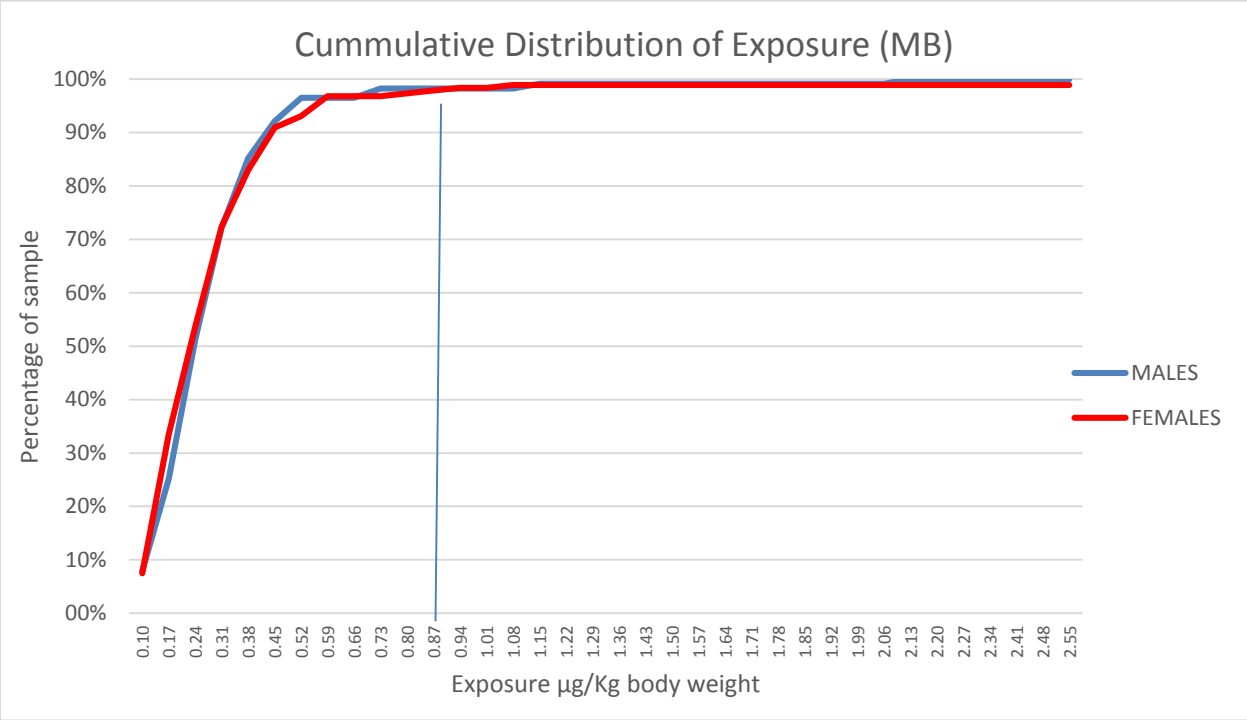
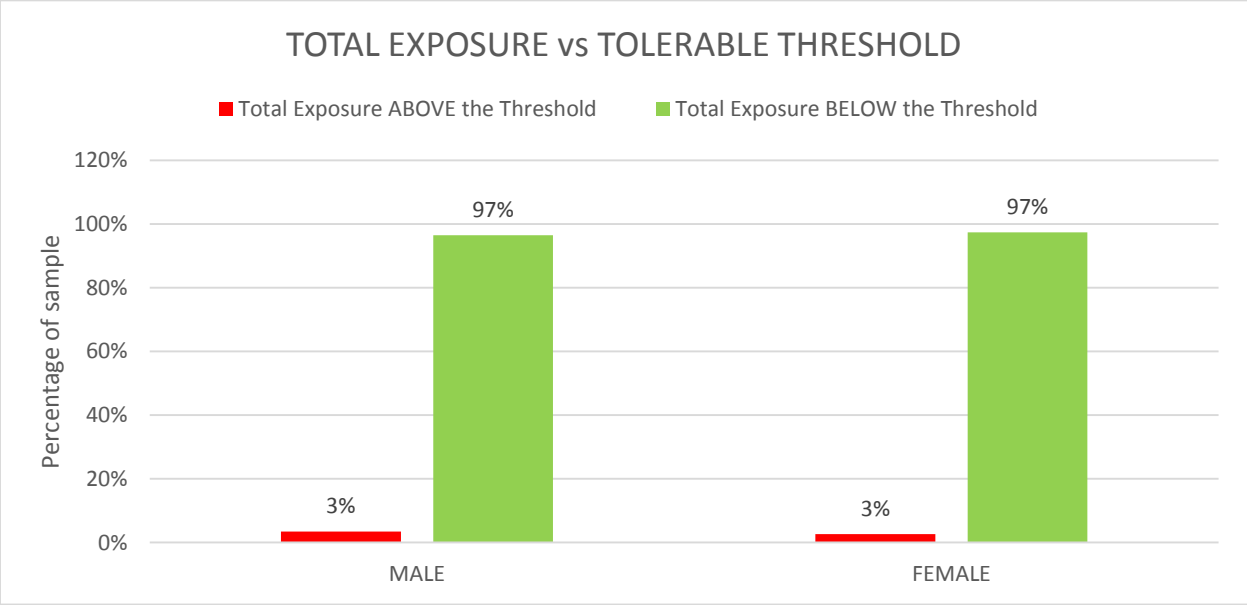


Figure 17: Proportion of the sample that exceed the tolerable threshold. Comparison across gender



Worksheet **AREA**

ImproRisk is using built[5] in formulas to identify the number and name of areas that are included in the *Subject* worksheet and creates a table of aggregated statistics (Figure 18). ImproRisk can handle up to 20 different Areas.

ImproRisk also produces comparative graphs for the mean exposure across area (Figure 19), and proportion of exceeding the tolerable threshold across the AREA. (Figure 20)

Figure 18: Statistics of the mean Exposure in the MB scenario by AREA of the individual. Q1: 25th percentile, Q3: 75th percentile, IQR: Interquartile range.95% error is the error (±) in the mean exposure. The error is 1.96 times the Standard Error (SE) of the estimated mean exposure

Total N=320	Area_1	Area_5	Area_2	Area_3	Area_4	Area_6
N	117	6	45	71	49	32
Min	0,05	0,27	0,19	0,14	0,14	0,08
Max	4,88	1,14	1,11	3,07	2,12	0,65
Q1	0,23	0,28	0,27	0,26	0,33	0,19
Median	0,32	0,29	0,35	0,38	0,48	0,28
Q3	0,63	0,43	0,50	0,49	0,63	0,39
IQR	0,40	0,15	0,23	0,23	0,30	0,20
Mean Exposure	0,78	0,46	0,41	0,50	0,60	0,30
St. Deviation of Exposure	1,10	0,34	0,19	0,49	0,42	0,16
St.Error	0,10	0,14	0,03	0,06	0,06	0,03
Lower Limit	0,58	0,18	0,35	0,38	0,48	0,24
Upper Limit	0,98	0,73	0,47	0,61	0,72	0,35
95% error	0,20	0,27	0,06	0,11	0,12	0,06
TOTAL EXPOSURE vs TOLERABLE THRESHOLD	Area_1	Area_5	Area_2	Area_3	Area_4	Area_6
Number of persons exposed above the Threshold	30	1	6	9	13	2
Total Exposure ABOVE the Threshold	26%	17%	13%	13%	27%	6%
Total Exposure BELOW the Threshold	74%	83%	87%	87%	73%	94%

[5] INDEX(AREA;MATCH(0;INDEX(COUNTIF($C$5:C5;AREA);0;0);0))

Figure 19: Mean Exposure in the MB scenario by AREA of the individual. The error lines on the bars represent the lower and upper limits for the 95% confidence interval of the true mean exposure.

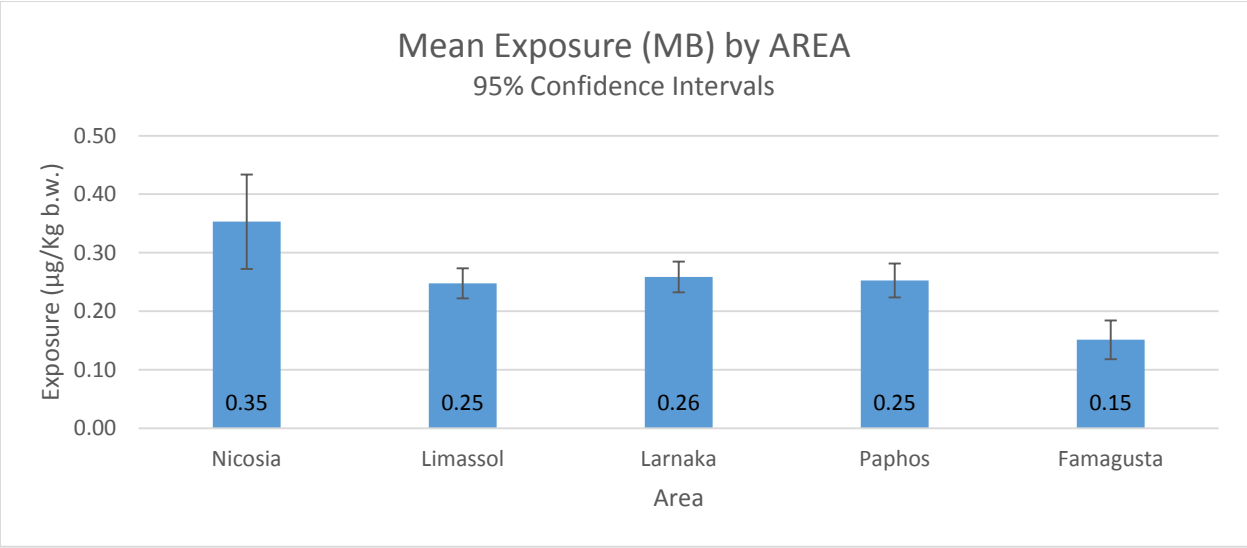
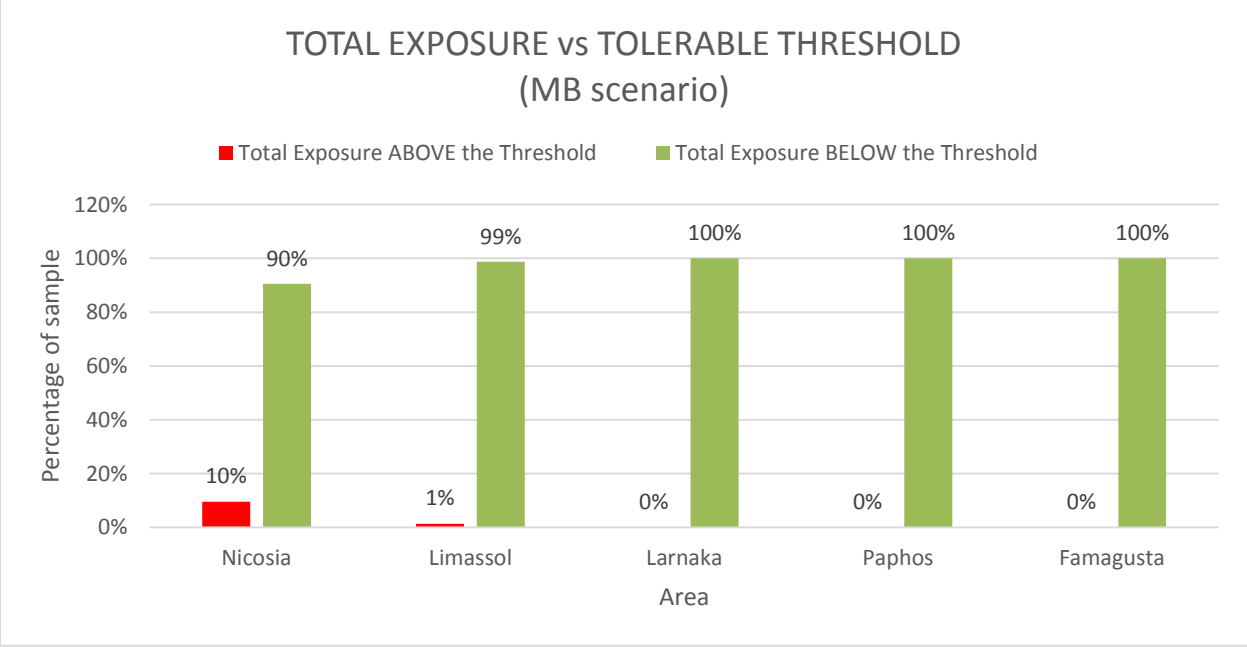


Figure 20: TOTAL EXPOSURE vs TOLERABLE THRESHOLD (MB scenario). The green bars represent the proportion of the sample that is "safe" i.e. below the tolerable threshold of exposure to the toxic The red bars represent the proportion of the sample which exceeds the tolerable threshold.

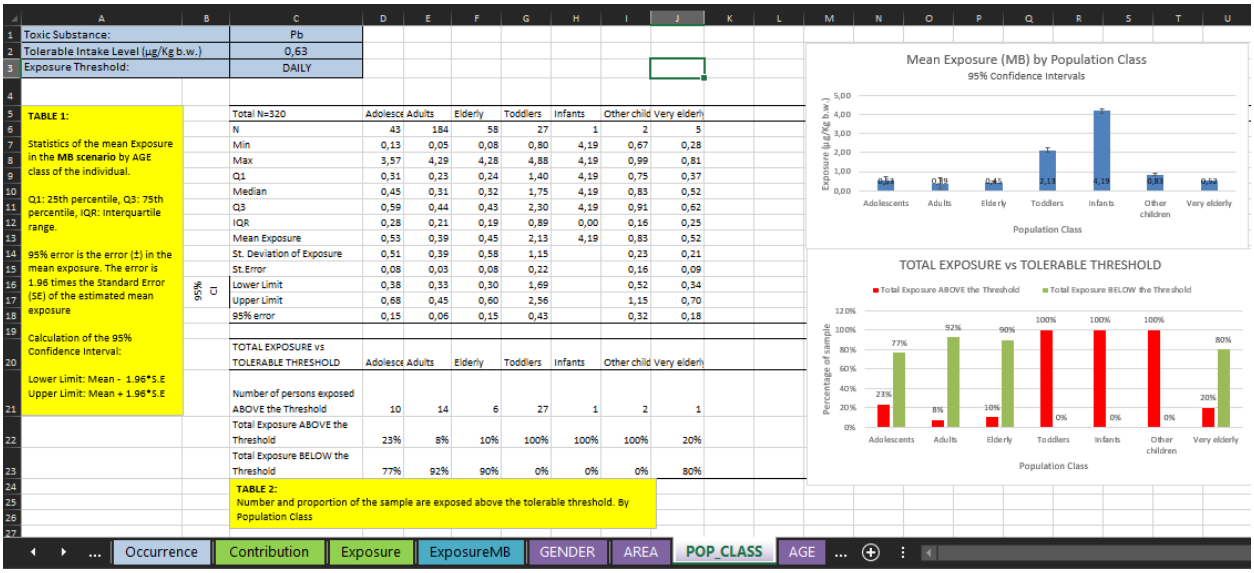


Worksheet **POP_CLASS**

ImproRisk is using built in formulas to identify the number and the name of Population Classes that are included in the **Subject** worksheet and creates a table of aggregated statistics (Figure 21). ImproRisk can handle up to 20 different Population Classes.

ImproRisk also produces comparative graphs for the mean exposure across Population Class and the proportion of the sample that exceeds the tolerable threshold across Population Class.

Figure 21: Screenshot of the worksheet POP_CLASS



Toxic Substance:	Pb
Tolerable Intake Level (µg/Kg b.w.)	0,63
Exposure Threshold:	DAILY

TABLE 1:

Statistics of the mean Exposure in the **MB scenario** by AGE class of the individual.

Q1: 25th percentile, Q3: 75th percentile, IQR: Interquartile range.

95% error is the error (±) in the mean exposure. The error is 1.96 times the Standard Error (SE) of the estimated mean exposure

Calculation of the 95% Confidence Interval:

Lower Limit: Mean - 1.96*S.E
Upper Limit: Mean + 1.96*S.E

Total N=320	Adolescents	Adults	Elderly	Toddlers	Infants	Other child	Very elderly
N	43	184	58	27	1	2	5
Min	0,13	0,05	0,08	0,80	4,19	0,67	0,28
Max	3,57	4,29	4,28	4,88	4,19	0,99	0,81
Q1	0,31	0,23	0,24	1,40	4,19	0,75	0,37
Median	0,45	0,31	0,32	1,75	4,19	0,83	0,52
Q3	0,59	0,44	0,43	2,30	4,19	0,91	0,62
IQR	0,28	0,21	0,19	0,89	0,00	0,16	0,25
Mean Exposure	0,53	0,39	0,45	2,13	4,19	0,83	0,52
St. Deviation of Exposure	0,51	0,39	0,58	1,15		0,23	0,21
St.Error	0,08	0,03	0,08	0,22		0,16	0,09
95% CI Lower Limit	0,38	0,33	0,30	1,69		0,52	0,34
95% CI Upper Limit	0,68	0,45	0,60	2,56		1,15	0,70
95% error	0,15	0,06	0,15	0,43		0,32	0,18

TOTAL EXPOSURE vs TOLERABLE THRESHOLD	Adolescents	Adults	Elderly	Toddlers	Infants	Other child	Very elderly
Number of persons exposed ABOVE the Threshold	10	14	6	27	1	2	1
Total Exposure ABOVE the Threshold	23%	8%	10%	100%	100%	100%	20%
Total Exposure BELOW the Threshold	77%	92%	90%	0%	0%	0%	80%

TABLE 2:
Number and proportion of the sample are exposed above the tolerable threshold. By Population Class

Worksheet **COMPREHENSIVE**

In this worksheet, one can calculate the mean exposure based on mean consumption data rather than consumption data at the individual level.

Mean consumption data can be retrieved (manually) via the EFSA data warehouse[6] (Figure 23) where lies the EFSA Comprehensive European Food Consumption Database. The Comprehensive Food Consumption Database is a source of information on food consumption across the European Union (EU). It contains detailed data for a number of EU countries.

EFSA published the "Guidance in the use of the EFSA Comprehensive European Food Consumption Database in Exposure Assessment" found in http://www.efsa.europa.eu/en/efsajournal/pub/2097.

One can retrieve the data via the EFSA data warehouse or alternatively can download the full datasets (in excel format .xlsx) from the Comprehensive data webpage[7]. The researcher can select from different datasets (chronic, acute, consumer based, etc) (see Figure 24). The excel files contain several worksheets each containing information down to specific FoodEx levels. Note that *ImproRisk* utilises the FoodEx level 2 categories.

How to use the worksheet

The food categories down to Level 2 are pre-installed (Columns A,B) (Figure 22). The user must fill in the Mean Consumption data he/she retrieved from the EFSA comprehensive dataset to columns C,D,E.

Do not copy-paste the full dataset you have retrieved from EFSA! Not all countries have information on all 160 food categories of FoodEx level 2! The ImproRisk worksheet allows for input of data for 161 food categories.

The Mean Occurrence is automatically retrieved from the Occurrence worksheet within ImproRisk.

The Exposure is then calculated via Excel formulas.

[6] https://dwh.efsa.europa.eu/bi/asp/Main.aspx?rwtrep=001
[7] http://www.efsa.europa.eu/en/food-consumption/comprehensive-database

Figure 22 Screenshot of the COMPREHENSIVE worksheet

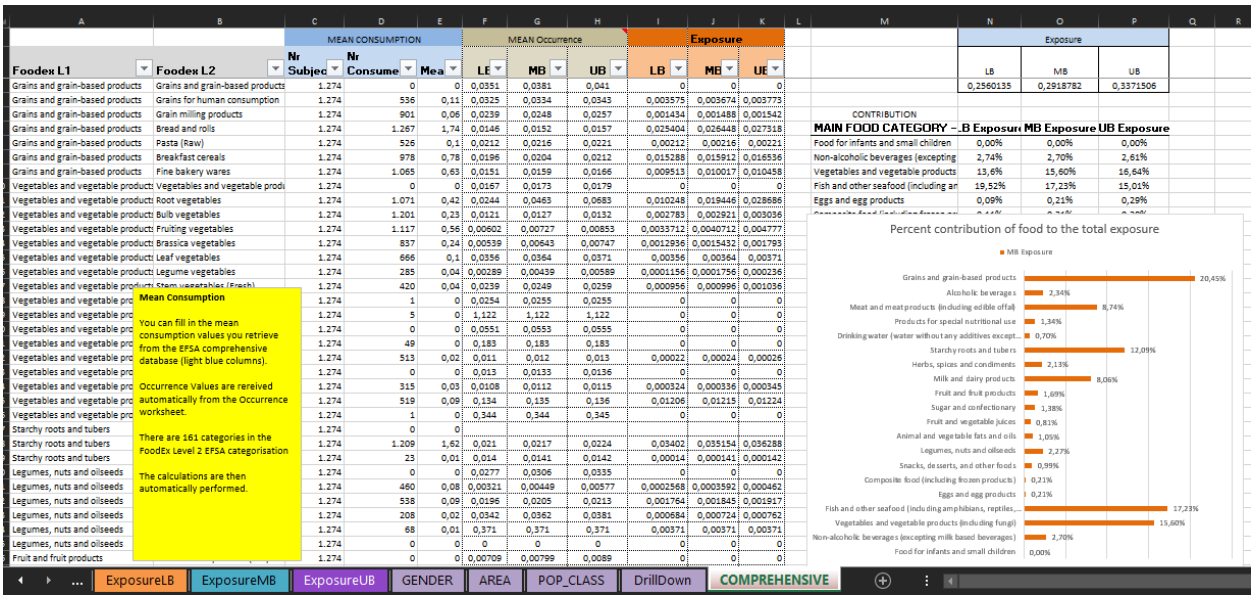


Figure 23 The EFSA Data Warehouse

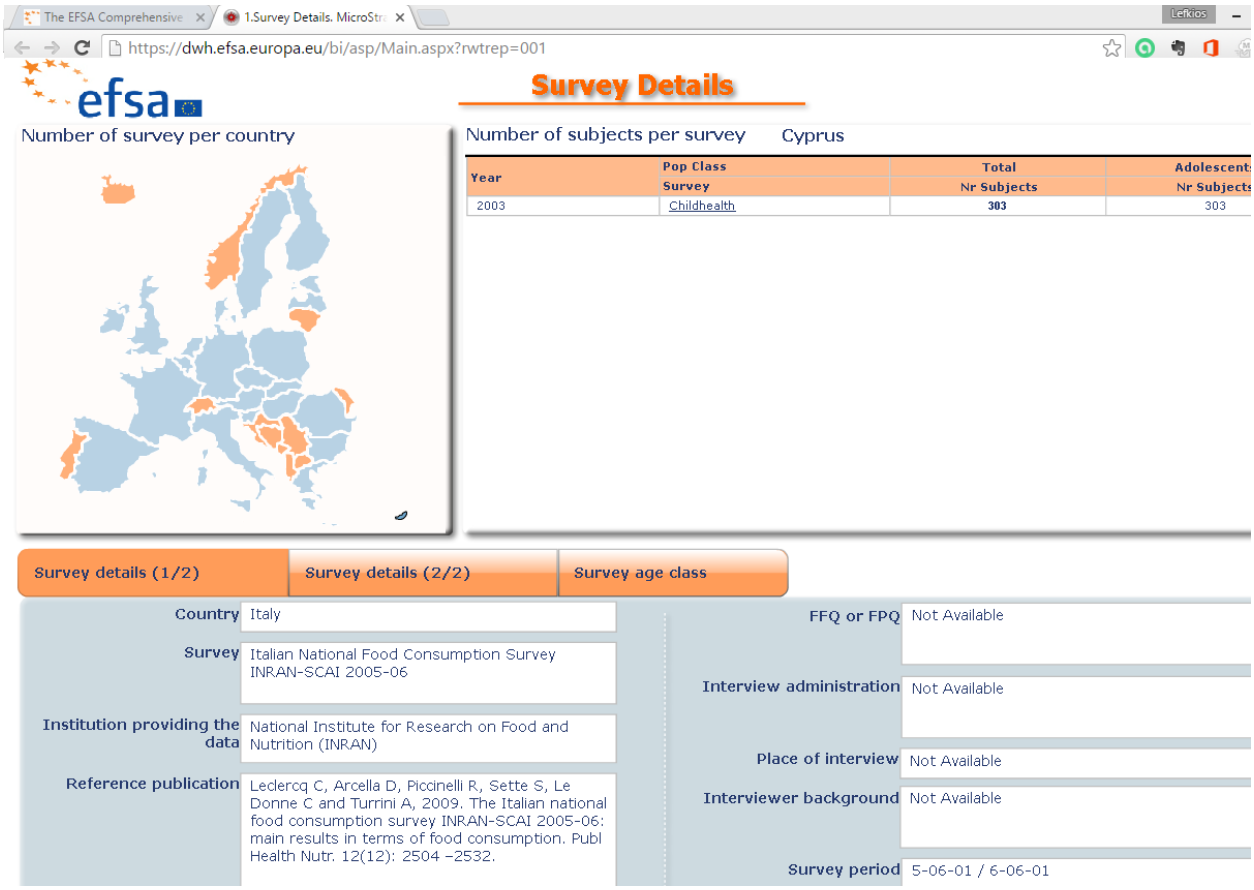
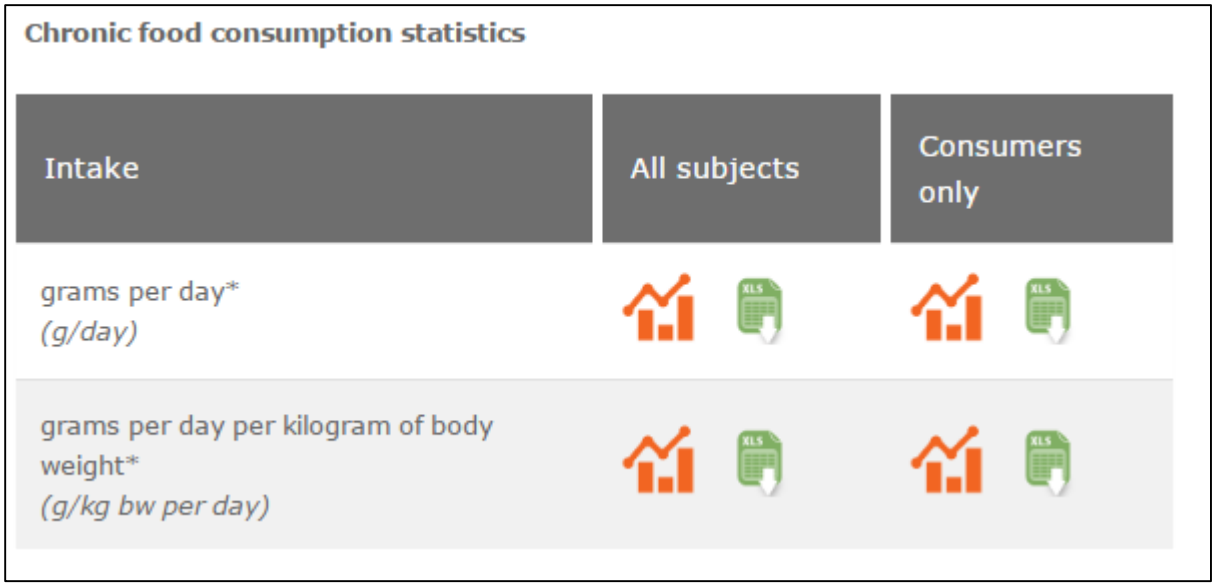


Figure 24 Screenshot for the EFSA Comprehensive European Food Consumption Database (http://www.efsa.europa.eu/en/food-consumption/comprehensive-database)



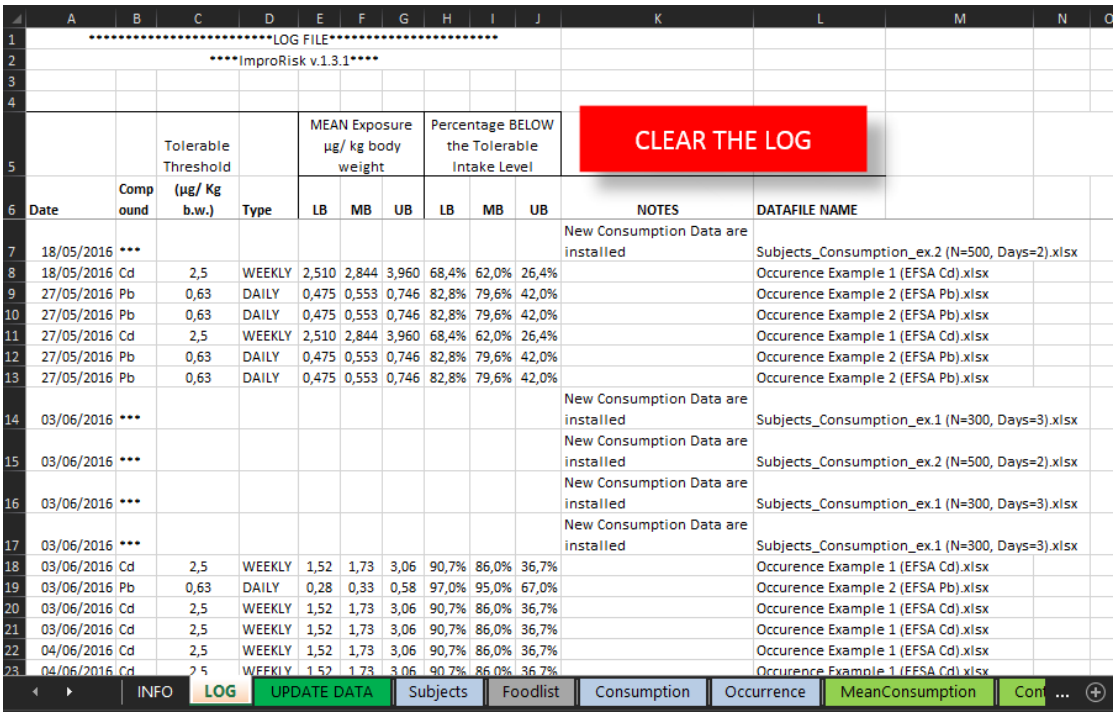
Chronic food consumption statistics

Intake	All subjects	Consumers only
grams per day* *(g/day)*		
grams per day per kilogram of body weight* *(g/kg bw per day)*		

LOG worksheet

This worksheet keeps a log of all the Exposure Assessments that are performed with ImproRisk 1.3.

Figure 25: Screenshot of the LOG worksheet



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	••••••••••••••••••••••••LOG FILE••••••••••••••••••••••••													
2	••••ImproRisk v.1.3.1••••													
3														
4														
5			Tolerable Threshold		MEAN Exposure µg/ kg body weight			Percentage BELOW the Tolerable Intake Level			CLEAR THE LOG			
6	Date	Compound	(µg/ Kg b.w.)	Type	LB	MB	UB	LB	MB	UB	NOTES	DATAFILE NAME		
7	18/05/2016	***									New Consumption Data are installed	Subjects_Consumption_ex.2 (N=500, Days=2).xlsx		
8	18/05/2016	Cd	2,5	WEEKLY	2,510	2,844	3,960	68,4%	62,0%	26,4%		Occurence Example 1 (EFSA Cd).xlsx		
9	27/05/2016	Pb	0,63	DAILY	0,475	0,553	0,746	82,8%	79,6%	42,0%		Occurence Example 2 (EFSA Pb).xlsx		
10	27/05/2016	Pb	0,63	DAILY	0,475	0,553	0,746	82,8%	79,6%	42,0%		Occurence Example 2 (EFSA Pb).xlsx		
11	27/05/2016	Cd	2,5	WEEKLY	2,510	2,844	3,960	68,4%	62,0%	26,4%		Occurence Example 1 (EFSA Cd).xlsx		
12	27/05/2016	Pb	0,63	DAILY	0,475	0,553	0,746	82,8%	79,6%	42,0%		Occurence Example 2 (EFSA Pb).xlsx		
13	27/05/2016	Pb	0,63	DAILY	0,475	0,553	0,746	82,8%	79,6%	42,0%		Occurence Example 2 (EFSA Pb).xlsx		
14	03/06/2016	***									New Consumption Data are installed	Subjects_Consumption_ex.1 (N=300, Days=3).xlsx		
15	03/06/2016	***									New Consumption Data are installed	Subjects_Consumption_ex.2 (N=500, Days=2).xlsx		
16	03/06/2016	***									New Consumption Data are installed	Subjects_Consumption_ex.1 (N=300, Days=3).xlsx		
17	03/06/2016	***									New Consumption Data are installed	Subjects_Consumption_ex.1 (N=300, Days=3).xlsx		
18	03/06/2016	Cd	2,5	WEEKLY	1,52	1,73	3,06	90,7%	86,0%	36,7%		Occurence Example 1 (EFSA Cd).xlsx		
19	03/06/2016	Pb	0,63	DAILY	0,28	0,33	0,58	97,0%	95,0%	67,0%		Occurence Example 2 (EFSA Pb).xlsx		
20	03/06/2016	Cd	2,5	WEEKLY	1,52	1,73	3,06	90,7%	86,0%	36,7%		Occurence Example 1 (EFSA Cd).xlsx		
21	03/06/2016	Cd	2,5	WEEKLY	1,52	1,73	3,06	90,7%	86,0%	36,7%		Occurence Example 1 (EFSA Cd).xlsx		
22	04/06/2016	Cd	2,5	WEEKLY	1,52	1,73	3,06	90,7%	86,0%	36,7%		Occurence Example 1 (EFSA Cd).xlsx		
23	04/06/2016	Cd	2,5	WEEKLY	1,52	1,73	3,06	90,7%	86,0%	36,7%		Occurence Example 1 (EFSA Cd).xlsx		

INFO | LOG | UPDATE DATA | Subjects | Foodlist | Consumption | Occurrence | MeanConsumption | Con ...

Fields in the Log file

Date: The date that the Exposure Assessment was performed.

Compound: The toxic substance that was investigated.

Tolerable Threshold: The tolerable threshold.

Type: The type of the tolerable threshold (DAILY or WEEKLY).

Mean Exposure: LB, MB, UB: The mean exposure for the LB, MB and UB scenarios in micrograms per body Kg body weight (µg/ Kg b.w.)

Percentage BELOW the Tolerable Intake Level: for the LB, MB and UB scenario.

Also, whenever new consumption or occurrence data are installed, then the file name is stored for quick reference.

Updating the Occurrence and Consumption

ImproRisk has the capability to change the Occurrence data as well as the Subject and Consumption data.

Occurrence

The occurrences, when in need to be updated or changed, **it must be done with a template that is provided** (Occurrence_Template.xlsx). The occurrences already in the ImproRisk model cannot be changed as the worksheet is locked for editing.

The template has the same structure as with the **Occurrence** worksheet in the ImproRisk.

The current version of ImproRisk performs exposure assessments on the average (mean) occurrences in the LB, MB, and UB scenario. The user can add occurrence values to the template only on columns H, I and J as seen in Figure 26.

Then three parameters must be defined:

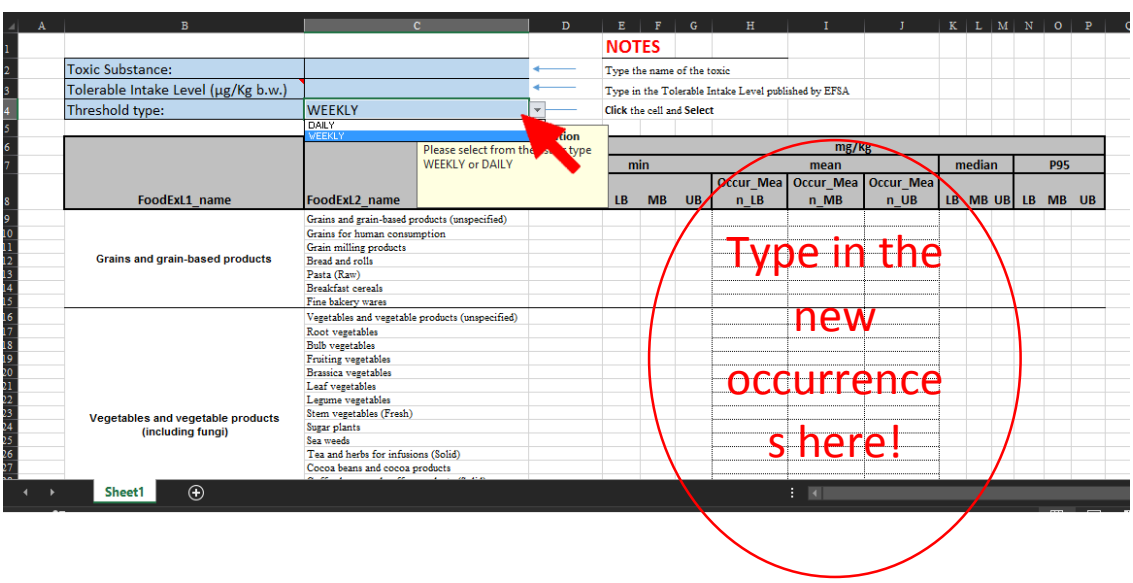
1. The name of the toxic substance; Enter the name in cell C2

2. The tolerable intake level in µg/Kg of body weight as provided by EFSA; Enter the value in cell C3

3. The threshold type of the tolerable intake level; i.e. is it a DAILY or WEEKLY threshold

Note that the template name (i.e. file name) is not important. Name the resulting file appropriately for easier reference. **What is of importance, it is the name of the worksheet within the template that needs to be named "Sheet1".**

Figure 26: Template for updating the occurrence data



When the template is saved, then use the UPDATE DATA worksheet, where VBA code lies for updating occurrence data (see *Figure 27*)

Steps:

1. Go to worksheet UPDATE DATA in the ImproRisk model

2. Click on the BLUE button to select an occurrence file

   The VBA code will ask the user to locate and select the file in the local disk

3. Navigate to the file that contains the occurrence file and click "Open"

The model will notify the user that a) Data are installed, b) Calculations are being performed and c) LOG file is being updated

4.. Wait for the model to until a message appears informing that "All Done!" (see Figure 28)

ALL the calculations and reports are automatically performed! Navigate to the other worksheets for the reporting.

Figure 27: Screenshot of the worksheet UPDATE DATA

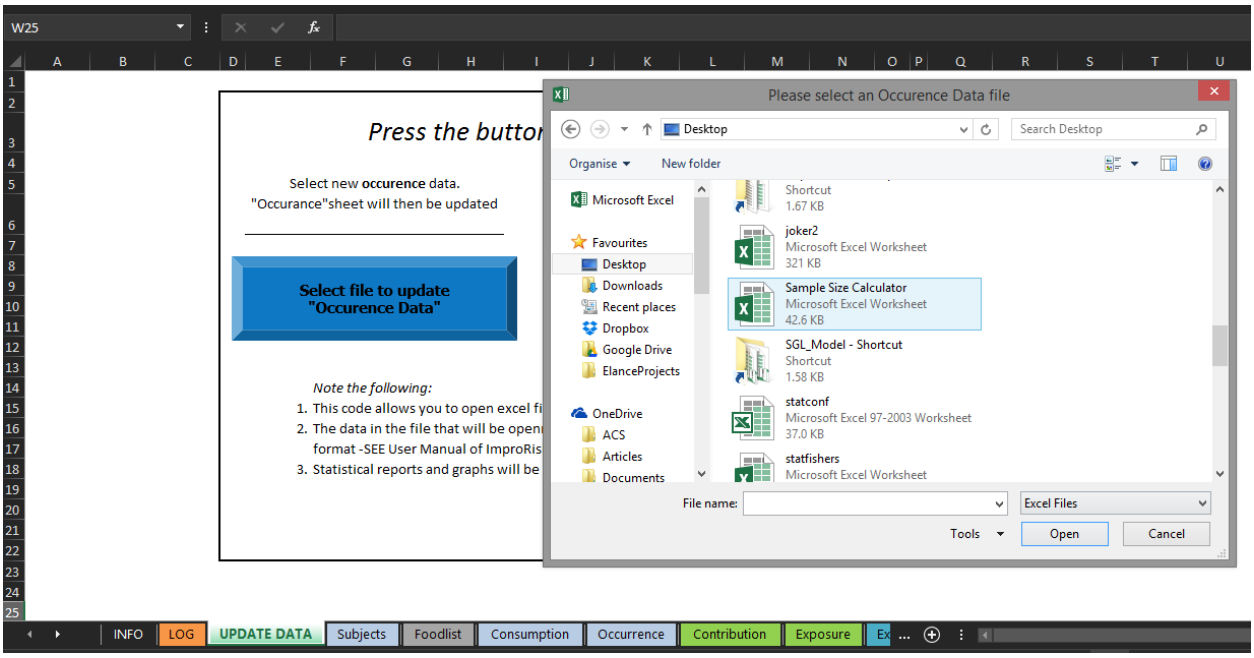
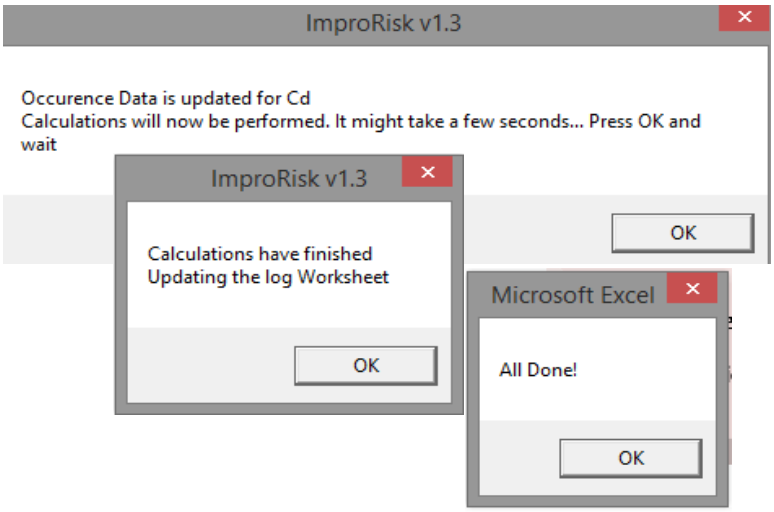


Figure 28: Notifications when updating the Occurrence data



Consumption

The Subject or the Consumption data, when in need to be updated or changed, **it must be done with a template that is provided** (Subjects_Consumption_Template.xlsx). The subject and consumption information already in the ImproRisk model cannot be changed as the worksheets are locked for editing.

There are two worksheets in the template:

<u>Worksheet "Subjects"</u>

This worksheet will contain the participants' demographic characteristics (See Figure 31)

When filling in the demographics, use...

1.MALE, FEMALE for the Gender

2.Numerical values for AGE and WEIGHT

3.Text values for AREA and POP_CLASS.

4. SUBJECTID must be a unique identifier; either numeric or a combination of text and number

Moreover, it is recommended to use the EFSA population classes approach when POP_CLASS column (Figure 29)

!!DO NOT change this worksheet name

Save the file with an appropriate name that is easily recognised by the user

Figure 29: Population Class by EFSA

POP_CLASS -	AGE
Infants	<1
Toddlers	1≤3
Other children	3≤10
Adolescents	10≤18
Adults	18≤65
Elderly	65≤75
Very elderly	≥75

**EFSA, 2011b*

Worksheets "Consumption"

This worksheet will contain the participants' food consumption occasions (see Figure 32)

Please note the following:

1. AMOUNTOFFOOD should be in grams

2. The name of the food (column E) must be in accordance to the FoodEx Level 4 categorisation by EFSA.

3. DAY contains the day of the food consumption occasion. **Enter a numerical value (1,2,3 etc.)**

4. In column AA, type in the sequential number (i.e. 1,2,3,4,...)

   In excel this is easily done by writing 1 and 2 at the first cells (A2 and A3), select the two cells and then drag down until the last food consumption occasion recorded. (see Figure 30)

!!DO NOT change this worksheet name

Save the file with an appropriate name that is easily recognised by the user

Figure 30: Creating sequential numbers in Excel

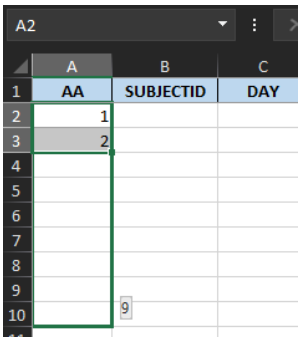


Figure 31: Template for updating the Subjects and Consumption data – Subjects worksheet

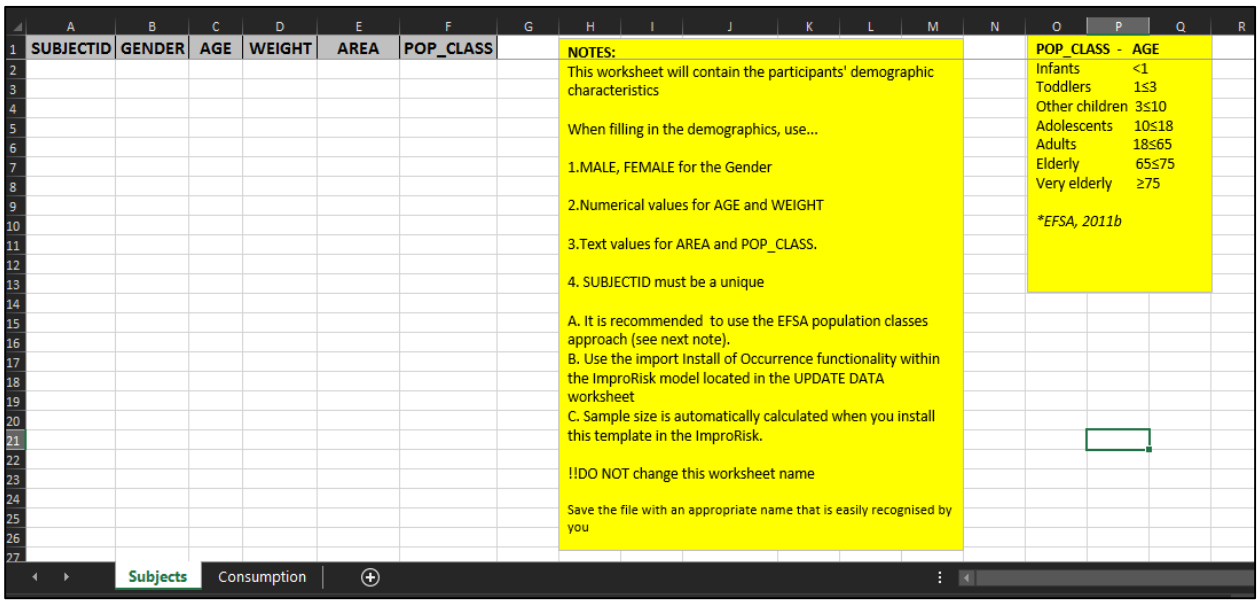
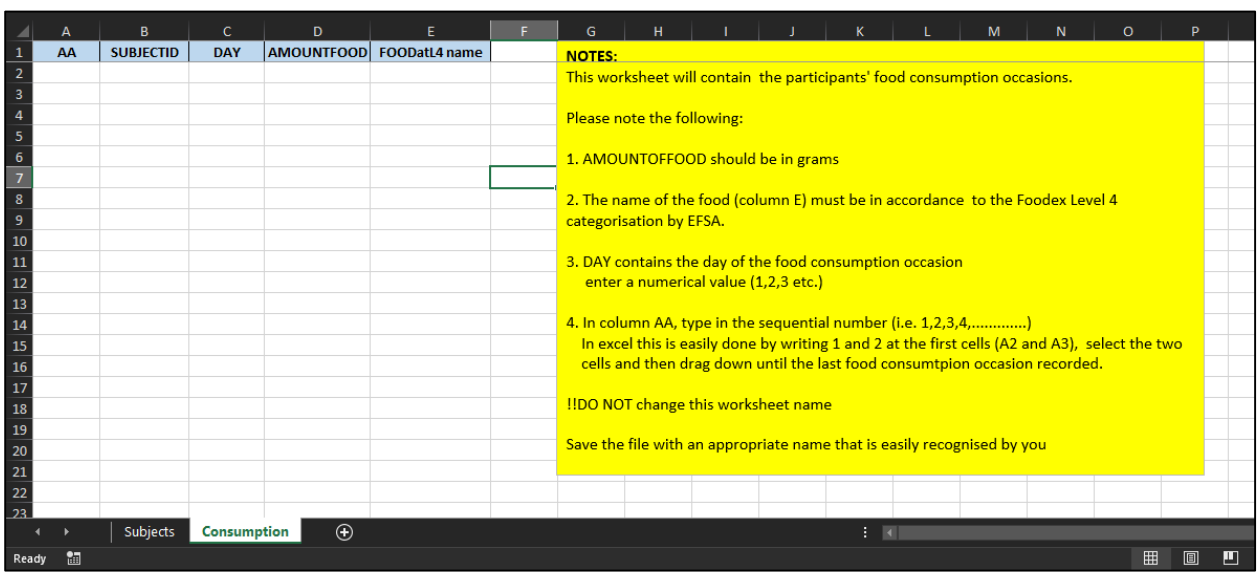


Figure 32: Template for updating the Subjects and Consumption data – Consumption worksheet



When the template is saved, then use the UPDATE DATA worksheet, where VBA code lies for updating occurrence data.

Steps:

1. Go to worksheet UPDATE DATA in the ImproRisk model

2. Click on the light PINK button for selecting a consumption file

The VBA code will ask the user to locate and select the file in the local disk

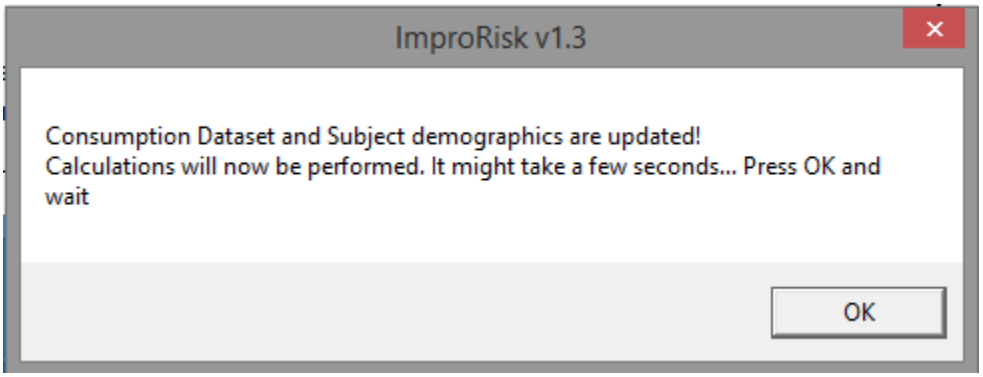
3. Navigate to the file that holds the consumption file and click "Open"

The model will notify the user that Consumption and Subject demographics are updated and that calculations will be executed.

4.. Wait for the model to until a message appears that informs that "All Done!" (see Figure 28)

ALL the calculations and reports are automatically performed! Navigate to the other worksheets for the reporting.

Figure 33: Notification the new Subject and Consumption data are updated



General Information

Automatic recalculation

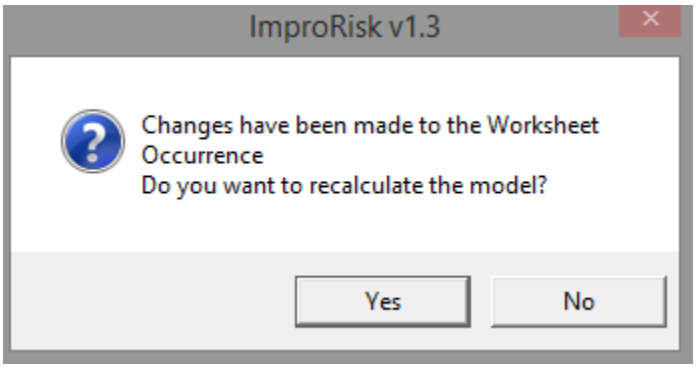
When new Occurrence or Consumption data are installed, VBA code recalculates all the formulas within ImproRisk. Automatic recalculation is disabled just after the calculations are performed.

However, the model recognizes when changes are made to certain worksheets (worksheets that contain information that have a direct impact on the exposure assessment), as soon as the user leaves the worksheet a message appears requesting permission to recalculate. Changes in Graphs do not count as changes, but changes in Pivot tables do count as changes.

In any case, any user who wishes to recalculate *the active worksheet*, this can be done when pressing *Shift-F9*. For recalculating *the entire model*, it can be done with F9.

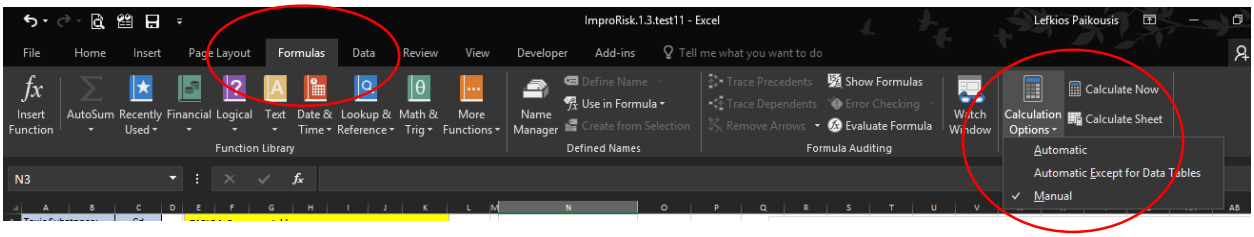
If the user wants to activate Automatic recalculation at any time (though not suggested since calculations may take significant amount of time depending on the subjects and food consumption occasions[8]), go to Formulas in the horizontal menu bar(ribbon) and go to Calculation Options (Figure 35).

Figure 34: Message when changes are detected in the ImproRisk model



[8] Currently it takes approximately 12 seconds to do the calculations based on 800 participants and 20,000 food consumption occasions

Figure 35: Automatic and Manual recalculation



Locked workbook and worksheet

ImproRisk and the worksheets are "locked for editing" meaning that the user cannot perform certain operations within the worksheets.

<u>The user cannot</u>:

1. Delete, move or change the name of the Worksheets
2. Insert, delete, move or change the columns within the worksheets
3. Type in data in the cells. All the cells are locked

   This means that if the user wants to update the Occurrence or Consumption database, the templates must be used and installed via the UPDATE DATA worksheet.

<u>The user is allowed to:</u>

1. Resize the worksheet's Columns and Rows
2. Format all the cells or columns
3. Edit the graphs
4. Use the PivotTables
5. Use AutoFilter
6. Edit all the objects in the workbook (i.e. note, graphs etc.)

The VBA code is also password protected

The column titles are unlocked

Notes within the worksheet

Within the worksheets there are several yellow notes that can be moved around. To move them, just click inside the note, and then "drag" them by touching on one of the edges.

VBA code

The VBA code is password protected. Unauthorised attempts to break the password protection are prohibited.

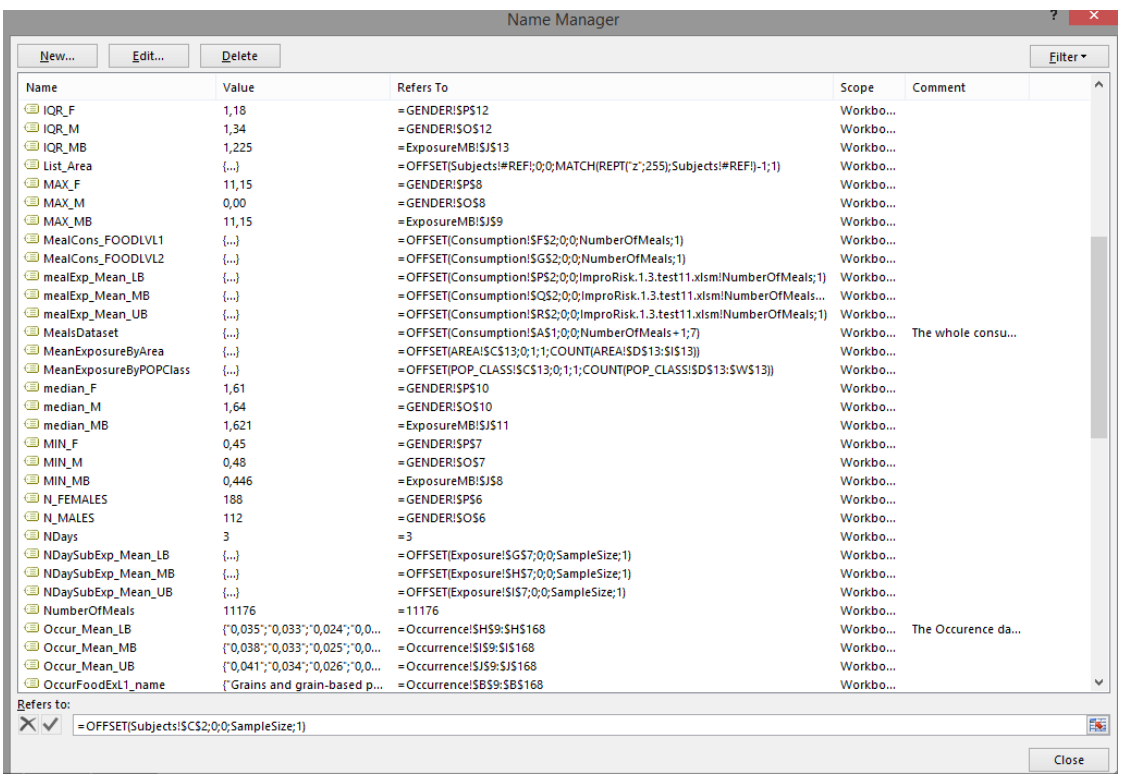
Named Ranges in Excel

Named Ranges are Excel names that refer to cells, a range of cells, a constant value, or a formula. ImproRisk uses those names in formulas, to replace values or cell references.

By using names, the formulas are much easier to read, understand and maintain.

i.e. named range *SUBJECTID* refers to all subjectID's in the sample, *NDays* holds the number of days of the food survey, *SubExp_MB* holds the total exposure (either Daily or Weekly) of each individual in the sample, etc.

Figure 36: Sample of Named Ranges in ImproRisk



Name	Value	Refers To	Scope	Comment
IQR_F	1,18	=GENDER!$P$12	Workbo...	
IQR_M	1,34	=GENDER!$O$12	Workbo...	
IQR_MB	1,225	=ExposureMB!$J$13	Workbo...	
List_Area	{...}	=OFFSET(Subjects!#REF!;0;0;MATCH(REPT("z";255);Subjects!#REF!)-1;1)	Workbo...	
MAX_F	11,15	=GENDER!$P$8	Workbo...	
MAX_M	0,00	=GENDER!$O$8	Workbo...	
MAX_MB	11,15	=ExposureMB!$J$9	Workbo...	
MealCons_FOODLVL1	{...}	=OFFSET(Consumption!$F$2;0;0;NumberOfMeals;1)	Workbo...	
MealCons_FOODLVL2	{...}	=OFFSET(Consumption!$G$2;0;0;NumberOfMeals;1)	Workbo...	
mealExp_Mean_LB	{...}	=OFFSET(Consumption!$P$2;0;0;ImproRisk.1.3.test11.xlsm!NumberOfMeals;1)	Workbo...	
mealExp_Mean_MB	{...}	=OFFSET(Consumption!$Q$2;0;0;ImproRisk.1.3.test11.xlsm!NumberOfMeals...	Workbo...	
mealExp_Mean_UB	{...}	=OFFSET(Consumption!$R$2;0;0;ImproRisk.1.3.test11.xlsm!NumberOfMeals;1)	Workbo...	
MealsDataset	{...}	=OFFSET(Consumption!$A$1;0;0;NumberOfMeals+1;7)	Workbo...	The whole consu...
MeanExposureByArea	{...}	=OFFSET(AREA!$C$13;0;1;1;COUNT(AREA!$D$13:$I$13))	Workbo...	
MeanExposureByPOPClass	{...}	=OFFSET(POP_CLASS!$C$13;0;1;1;COUNT(POP_CLASS!$D$13:$W$13))	Workbo...	
median_F	1,61	=GENDER!$P$10	Workbo...	
median_M	1,64	=GENDER!$O$10	Workbo...	
median_MB	1,621	=ExposureMB!$J$11	Workbo...	
MIN_F	0,45	=GENDER!$P$7	Workbo...	
MIN_M	0,48	=GENDER!$O$7	Workbo...	
MIN_MB	0,446	=ExposureMB!$J$8	Workbo...	
N_FEMALES	188	=GENDER!$P$6	Workbo...	
N_MALES	112	=GENDER!$O$6	Workbo...	
NDays	3	=3	Workbo...	
NDaySubExp_Mean_LB	{...}	=OFFSET(Exposure!$G$7;0;0;SampleSize;1)	Workbo...	
NDaySubExp_Mean_MB	{...}	=OFFSET(Exposure!$H$7;0;0;SampleSize;1)	Workbo...	
NDaySubExp_Mean_UB	{...}	=OFFSET(Exposure!$I$7;0;0;SampleSize;1)	Workbo...	
NumberOfMeals	11176	=11176	Workbo...	
Occur_Mean_LB	{"0,035";"0,033";"0,024";"0,0...	=Occurrence!$H$9:$H$168	Workbo...	The Occurence da...
Occur_Mean_MB	{"0,038";"0,033";"0,025";"0,0...	=Occurrence!$I$9:$I$168	Workbo...	
Occur_Mean_UB	{"0,041";"0,034";"0,026";"0,0...	=Occurrence!$J$9:$J$168	Workbo...	
OccurFoodExL1_name	{"Grains and grain-based p...	=Occurrence!$B$9:$B$168	Workbo...	

Statistical Terms

Cumulative Frequency

Cumulative means "how much so far"[9]. The total of a frequency and all frequencies so far in a frequency distribution. It is the 'running total' of frequencies.

To have cumulative totals, just add up the values as you go.

Figure 37: Cumulative frequency example. Thank you www.mathisfun.com

Cumulative percentage

The total of a percentage and all percentages so far in a frequency distribution. It is the 'running total' of percentages.

To have cumulative totals, just add up the values as you go.

Otherwise, it is the Cumulative frequency divided by the total number of subjects. Please see nice description of the cumulative percentage use and benefits by the state statistical service of Canada [10]

Quartiles

When the distribution of exposure is described in the **ExposureMB** or **GENDER** worksheets, summary statistics named Q1, Q3 and IQR are presented.

Q1 = 1st quartile or 25th percentile.

Median = 2nd quartile or 50th percentile

[9] https://www.mathsisfun.com/definitions/cumulative-frequency.html
[10] http://www.statcan.gc.ca/edu/power-pouvoir/ch10/5214864-eng.htm

Q1 = 3rd quartile or 75th percentile

IQR = Q3-Q1

e.g. if the Q1 of exposure is 1.19 µg/Kg b.w., this means 25% of the sample are exposed UP TO 1.19. In other words, 25% of the subjects have exposure below 1.19 µg/Kg b.w.

e.g. if the Q3 of exposure is 2.53, this means this means 75% of the sample are exposed UP TO 2.53 µg/Kg. In other words, 25% of the sample are exposed more than 2.53 µg/Kg b.w.

e.g. If the Median exposure is 1.64 µg/Kg b.w. then half of the sample is exposed less than 1.64 µg/Kg b.w. and half above 1.64 µg/Kg b.w.

Probability Distribution (histogram)

It is better known as the histogram[11].

A histogram is a graphical representation of the distribution of **exposure values** of all the subjects in the food survey grouped into ranges. It is similar to a Bar Graph, but in a Histogram each bar is for a range of data.

The entire range of exposure values is divided into a series of intervals—and then a count is obtained of how many exposure values (i.e. subjects) fall into each interval. The bins are specified as consecutive, non-overlapping intervals of the exposure values. The bins (intervals) must be adjacent, and they are of equal size. In ImproRisk the size of the bin is called "height". It is calculated via the Friedman-Diaconis rule[12] and it is very robust in practice (Freedman D. 1981, Scott D. 1992) when compared to Sturges' rule that is used by a variety of statistical software[13]

For example, in Figure 38, 21.7% of the sample has exposure between 1.19 and 1.56 µg/Kg b.w.

Cumulative Distribution

It is a graphical representation of the Cumulative Percentages (see *Cumulative percentage*). Put simply, it is the running total of the bars in a histogram. The end of the running total will be 100%.

It is a better visualisation when two groups are compared as to their exposure i.e. Male and Female. In Figure 39 the red line for Female is higher than Male. This means that more females are below[14] any exposure value than men are.

---

[11] https://en.wikipedia.org/wiki/Histogram

[12] https://en.wikipedia.org/wiki/Freedman%E2%80%93Diaconis_rule

[13] http://robjhyndman.com/papers/sturges.pdf

[14] Remember that the cumulative is a running total – it is the percentage UP to and including a particular exposure

In general, the higher the curve, the better off.

Figure 38: Example of a histogram

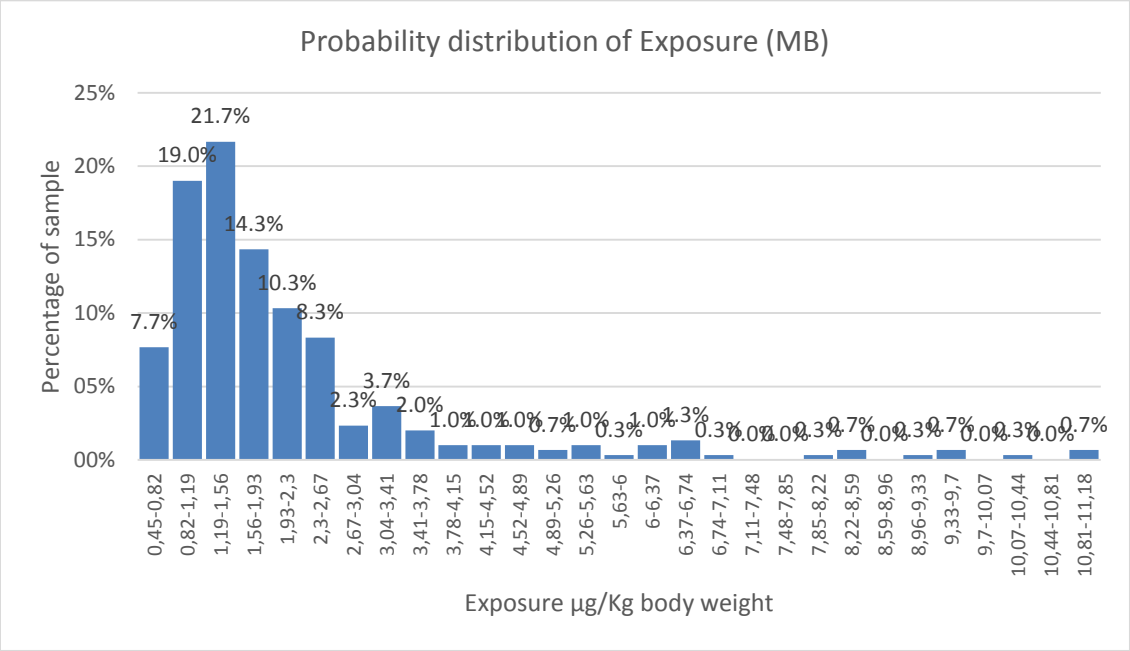
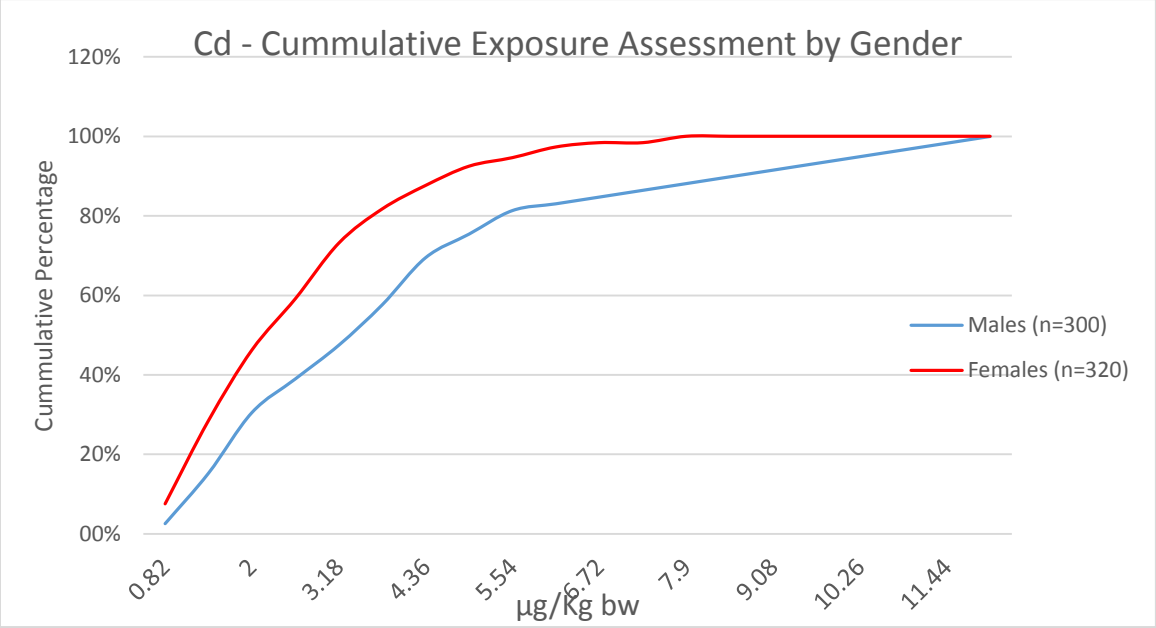


Figure 39: Cumulative probability across Gender



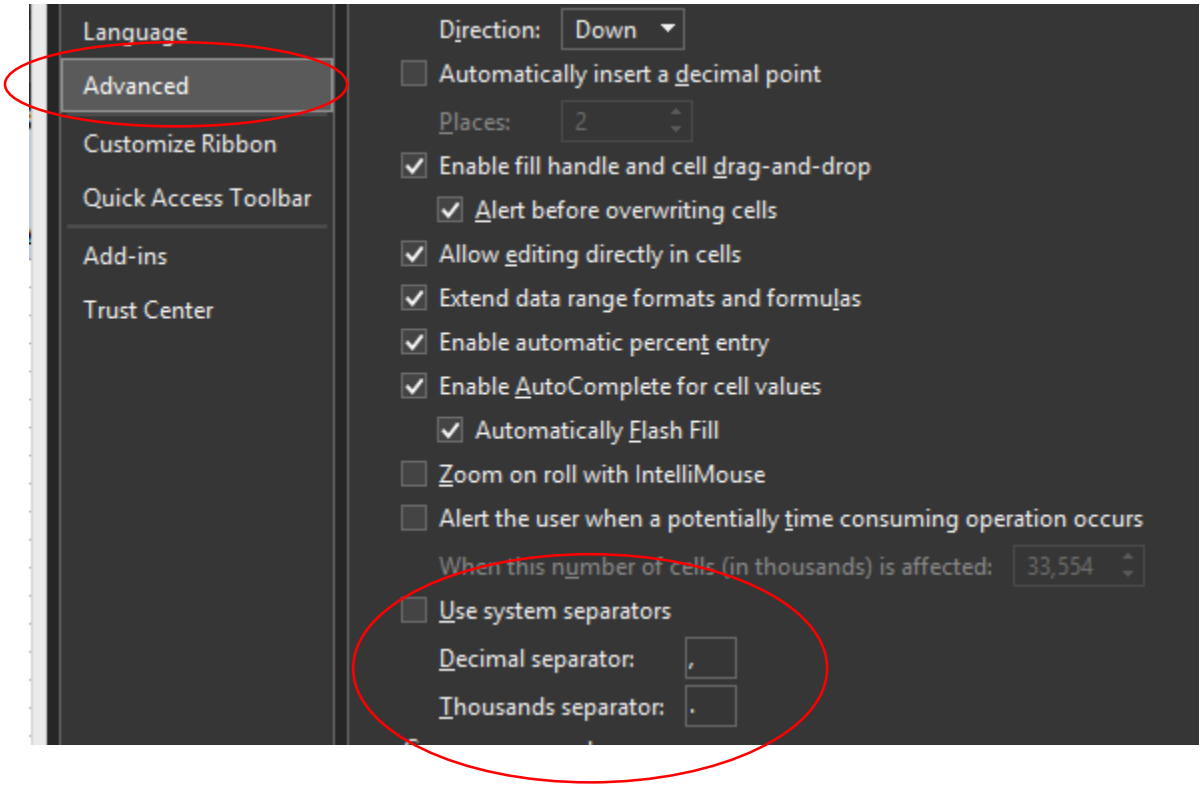
Decimal and Thousand separator

Whatever the decimal or thousand separators were during building the ImproRisk, when opening the model values will adjust to your local settings.

Should you wish to change the decimal and thousands separators in Excel go to..

File - > move down and click on ***Options***-> move down and click on ***Advanced***-> On the right panel look for the "Use system separators section" (Figure 40).

Figure 40: Changing the decimal and thousand separators in Excel



References

1. Freedman, David; Diaconis, Persi (December 1981). "On the histogram as a density estimator: L2 theory" Probability Theory and Related Fields (Heidelberg: Springer Berlin) 57 (4): 453–476.

2. Scott, D. (1992). Multivariate density estimation. New York: Wiley.

3.STATISTICS CANADA. Cumulative percentage. [online] Statcan.gc.ca. Available at: http://www.statcan.gc.ca/edu/power-pouvoir/ch10/5214864-eng.htm [Accessed 10 May 2016].

4. Cohen, J. (1969) Statistical Power Analysis for the Behavioral Sciences. NY: Academic Press

5. Coe, R. (2002). It's the Effect Size, Stupid. What effect size is and why it is important. Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England

6. University, Penn State. Comparing Two Independent Means. [online] STAT 200 Elementary Statistics. Available at: https://onlinecourses.science.psu.edu/stat200/node/60 [Accessed 5 May 2016].

7. European Food Safety Authority; Evaluation of the FoodEx, the food classification system applied to the development of the EFSA Comprehensive European Food Consumption Database. EFSA Journal 2011; 9(3):1970. http://www.efsa.europa.eu/sites/default/files/scientific_output/files/main_documents/1970.pdf

8. European Food Safety Authority; Use of the EFSA Comprehensive European Food Consumption Database in Exposure Assessment. EFSA Journal 2011;9(3):2097. http://www.efsa.europa.eu/sites/default/files/scientific_output/files/main_documents/2097.pdf

9. European Food Safety Authority; Cadmium dietary exposure in the European population. EFSA Journal 2012;10(1):2551 http://www.efsa.europa.eu/sites/default/files/scientific_output/files/main_documents/2551.pdf

10. European Food Safety Authority; Lead dietary exposure in the European population. EFSA Journal 2012;10(7):2831. [59 pp.] http://www.efsa.europa.eu/sites/default/files/scientific_output/files/main_documents/2831.pdf

Appendix A – Updates in ImproRisk 1.3.1

ImproRisk 1.3.1 contains four main changes with respect to version 1.3.

Probability and Cumulative distribution function for LB and UB exposure

ImproRisk now produces probability and cumulative distribution functions for both the LB (Lower Bound) and MB (Middle Bound) scenario of exposure.

The calculations are the same as in the MB scenario, but using individual exposure based on the LB (and UB) occurrence of the toxic under study.

The graphs are depicted in Orange for the LB scenario (Figure 41) and Purple for the UB scenario (Figure 42). Each scenario is analysed in separate worksheets (namely: *ExposureLB* and *ExposureUB*)

Figure 41 Probability distribution for the LB scenario

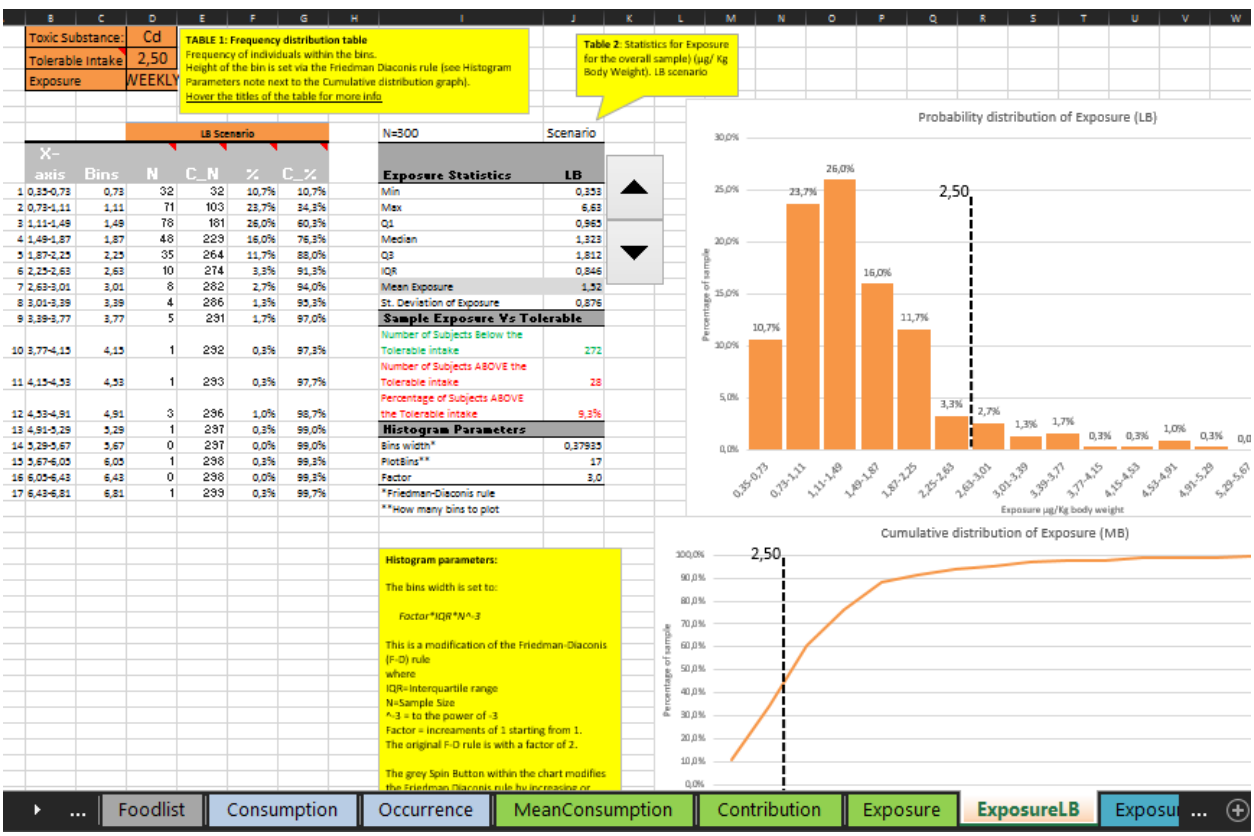
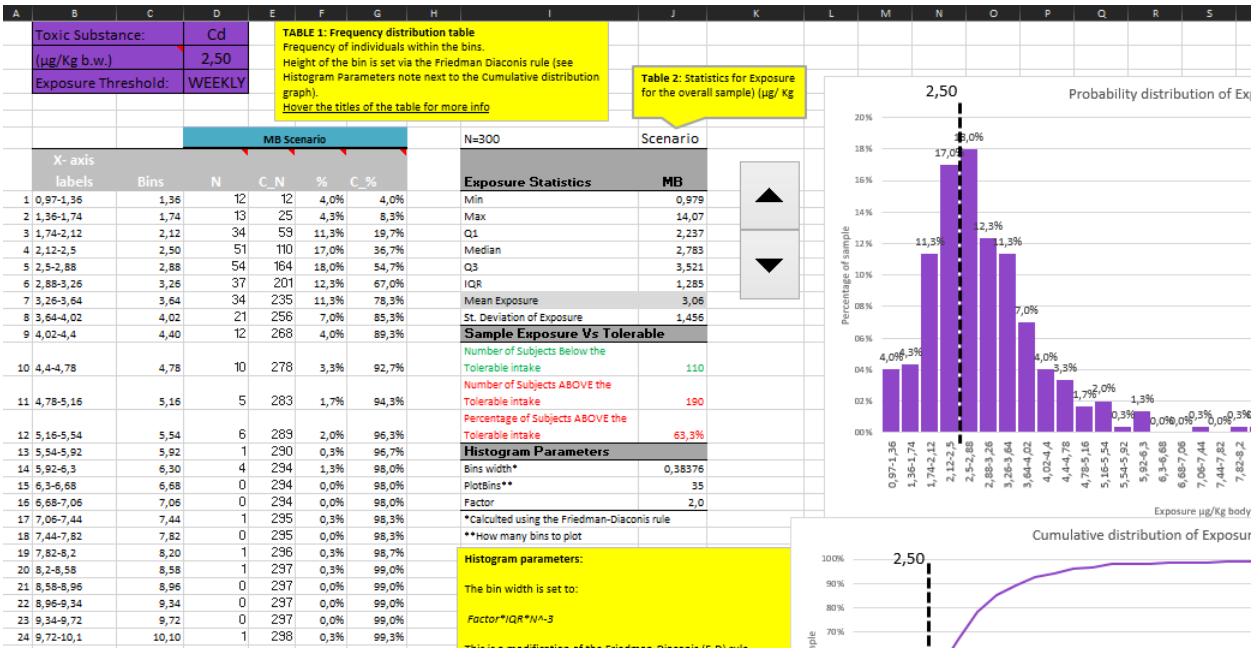


Figure 42 Probability distribution for the UB scenario



Toxic Substance:	Cd
(µg/Kg b.w.)	2,50
Exposure Threshold:	WEEKLY

TABLE 1: Frequency distribution table
Frequency of individuals within the bins.
Height of the bin is set via the Friedman Diaconis rule (see Histogram Parameters note next to the Cumulative distribution graph).
Hover the titles of the table for more info

Table 2: Statistics for Exposure for the overall sample) (µg/ Kg

	X- axis labels	Bins	MB Scenario N	C_N	%	C_%
1	0,97-1,36	1,36	12	12	4,0%	4,0%
2	1,36-1,74	1,74	13	25	4,3%	8,3%
3	1,74-2,12	2,12	34	59	11,3%	19,7%
4	2,12-2,5	2,50	51	110	17,0%	36,7%
5	2,5-2,88	2,88	54	164	18,0%	54,7%
6	2,88-3,26	3,26	37	201	12,3%	67,0%
7	3,26-3,64	3,64	34	235	11,3%	78,3%
8	3,64-4,02	4,02	21	256	7,0%	85,3%
9	4,02-4,4	4,40	12	268	4,0%	89,3%
10	4,4-4,78	4,78	10	278	3,3%	92,7%
11	4,78-5,16	5,16	5	283	1,7%	94,3%
12	5,16-5,54	5,54	6	289	2,0%	96,3%
13	5,54-5,92	5,92	1	290	0,3%	96,7%
14	5,92-6,3	6,30	4	294	1,3%	98,0%
15	6,3-6,68	6,68	0	294	0,0%	98,0%
16	6,68-7,06	7,06	0	294	0,0%	98,0%
17	7,06-7,44	7,44	1	295	0,3%	98,3%
18	7,44-7,82	7,82	0	295	0,0%	98,3%
19	7,82-8,2	8,20	1	296	0,3%	98,7%
20	8,2-8,58	8,58	1	297	0,3%	99,0%
21	8,58-8,96	8,96	0	297	0,0%	99,0%
22	8,96-9,34	9,34	0	297	0,0%	99,0%
23	9,34-9,72	9,72	0	297	0,0%	99,0%
24	9,72-10,1	10,10	1	298	0,3%	99,3%

N=300	Scenario
Exposure Statistics	**MB**
Min	0,979
Max	14,07
Q1	2,237
Median	2,783
Q3	3,521
IQR	1,285
Mean Exposure	3,06
St. Deviation of Exposure	1,456
Sample Exposure Vs Tolerable	
Number of Subjects Below the Tolerable intake	110
Number of Subjects ABOVE the Tolerable intake	190
Percentage of Subjects ABOVE the Tolerable intake	63,3%
Histogram Parameters	
Bins width*	0,38376
PlotBins**	35
Factor	2,0

*Calculted using the Friedman-Diaconis rule
**How many bins to plot

Histogram parameters:

The bin width is set to:

*Factor*IQR*N^-3*

Pop Up Menu for quick navigation across worksheets

Using VBA code, a Pop-Up Menu is created that enables the user to quickly navigate through the worksheets.

The Pop-Up Menu is activated by either of the following ways:

1. Press ***CTRL-m***

   The Pop-Up menu will appear. The user can select any worksheet and will be redirected to it.

2. ***Right-click*** on any cell and select "ImproRisk Pop-Up Menu"

   Then the Pop-Up menu will appear

Both ways can be applied at any time within any worksheet.

Figure 43 Pop Up Menu for navigating the ImproRisk worksheets (Right-click on any cell within any worksheet)

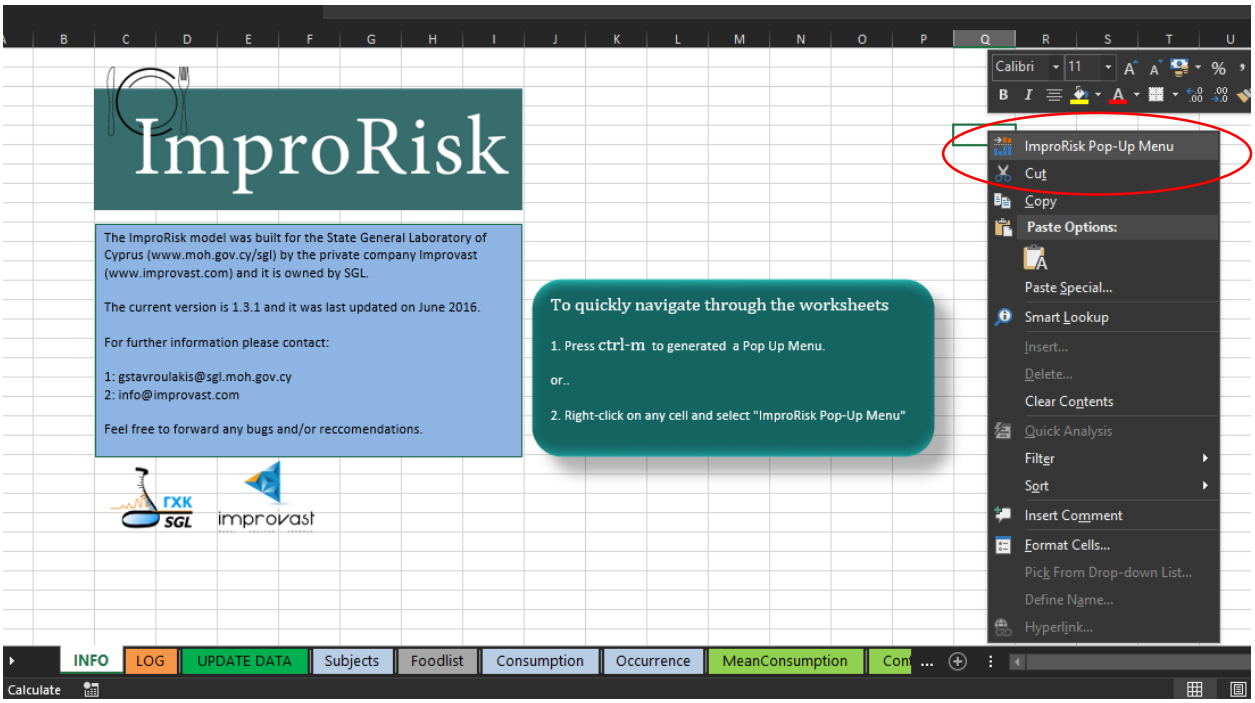
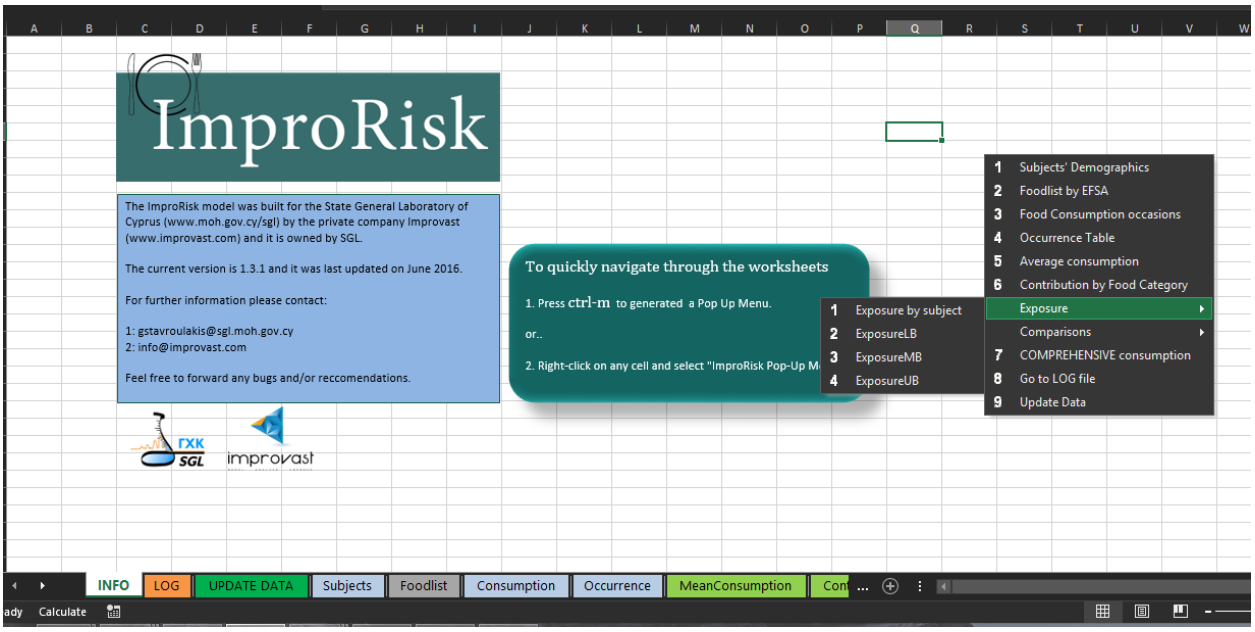


Figure 44 Pop Up Menu for navigating the ImproRisk worksheets (Press ctrl-m)



Separate worksheet for the Chronic Consumption calculation

A separate worksheet is created where the average daily consumption across each food category is calculated.

The tables for the mean consumption, used to be in the worksheet Exposure in the ImproRisk 1.3. Now a separate worksheet (*MeanConsumption*) presents the calculations, final tables, and graphical representation of the food category contribution to the mean consumption.

Note that the calculations are such so that the chronic consumption is calculated[15].

Optimisation of the VBA syntax

VBA syntax was optimised to avoid redundant actions therefore reducing calculation and execution speed.

[15] http://www.efsa.europa.eu/en/efsajournal/pub/2097