Weighting coefficients in the ImproRisk© model

The ImproRisk version 1.4 accommodates weighting coefficients to adjust the sample for non-representativeness within the population. Exposure statistics e.g. mean exposure and standard error, will now be adjusted to reflect the population from which the sample was taken from.

Each individual in the food survey sample will carry a "weight coefficient". This coefficient will be a number (>1) that indicates the number of persons in the population that this individual represents. If for example an individual's weight coefficient is 500, then that individual represents 500 persons in the population of the same population class.

Weight coefficients, or simply "weights" need to be calculated beforehand by the risk assessor. The weight coefficients will then be installed in ImproRisk within the *Subjects_Consumption* template provided (see

Subjects and Consumption template section). **In case there are no weight coefficients assigned, the weight for each individual is set to *"1"*** (i.e. subject represents only him/herself)

Typically, weight coefficients are calculated using (recent) census data of the population.

NOTE! ImproRisk uses frequency weights[1] and not reliability weights. The formulae for adjusted estimates derive unbiased estimates for frequency weights and not for reliability weights.

A typical way of calculating weights

Let's assume the user wants to adjust the sample by **Gender** and **Age Group**.

First, a (cross tabulation) table of real **population counts** in each Gender by Age combination needs to be calculated. This is where the census data will be used.

Assuming that the age group has 3 levels (Children, Adults, Elderly) and Gender has 2 levels (Males, Females), the cross tabulation gives a table of the total number of persons in each Gender by Age class, a total of 6 Classes (2X3=6).

The corresponding **sample counts** will be calculated from the Sample data again by tabulation.

The weight to be applied, **is the ratio of the Population counts to the Sample counts for the Class to which the person belongs**. This means that subjects that belong to the same "population" class will carry the same weighting coefficient, i.e. Adult men will all have the same "weight", adult women the same, etc.

In case the user wants to adjust for more than three population attributes (e.g. Gender, Age and Area), it is essential to note of the possibility (depending in the sample size and sampling strategy) that the sample crosstabulation may give many cells with no data or very small sample numbers. This would give very high weights to some individuals and lead to inaccurate results.

[1] https://en.wikipedia.org/wiki/Weighted_arithmetic_mean#Frequency_weights

A methodology for calculating weights for a non-representative sample is described on page 64 of the guide to the 2007 Australian National Children's Nutrition and Physical Activity Survey:

https://www.health.gov.au/internet/main/publishing.nsf/Content/589EFDBF5E7B916FCA257BF000211E08/$File/user-guide-v2.pdf

Example - Cyprus

An example is provided to calculate the weight coefficients for the Cyprus sample (n=300) to adjust for the population (N=706652) (Figure 1).

The crosstabulation for the sample counts (blue table) shows the distribution of gender and age group in the sample. The crosstabulation for the population counts (green table) shows the respective counts as extracted by tabulating the latest census data (CYSTAT, 2011)[2]

The pink table (Weights rounded), shows the resulting weight coefficients for the 6 classes

e.g. The weight coefficient for Male Adolescents will be 52862/19=2782

e.g. The weight coefficient for Male Adults will be 259221/77=3367

etc.

It can be derived that the female proportion in the sample is 62.7% (188/300) while the female proportion in the population is 51.4% (362869/706652). This means that females are overrepresented in the sample. This is now adjusted when the weight coefficient for females for all age groups, is lower than that of males.

Note that the weights do not need to be integers. In this example, we round the weights to the nearest integer for simplicity reasons.

Figure 1: Weight coefficients calculated for the Cyprus population for a sample of 300

SUBJECTID	GENDER	AGE	WEIGHT	AREA	POP_CLASS	WCOEFF
1001	FEMALE	12	81	Nicosia	Adolescents	2094
1002	FEMALE	15	43	Nicosia	Adolescents	2094
1003	FEMALE	62	69	Nicosia	Adults	2302
1004	FEMALE	63	65	Nicosia	Adults	2302
1005	FEMALE	64	72	Nicosia	Adults	2302
1006	FEMALE	65	71	Nicosia	Adults	2302
1007	FEMALE	29	57	Nicosia	Adults	2302
1008	FEMALE	65	64	Nicosia	Adults	2302
1009	MALE	60	62	Nicosia	Adults	3367
1010	MALE	50	104	Nicosia	Adults	3367
1011	MALE	58	69	Nicosia	Adults	3367
1012	MALE	45	59	Nicosia	Adults	3367
1013	MALE	70	58	Nicosia	Elderly	1981
1014	MALE	44	61	Nicosia	Adults	3367
1015	MALE	64	91	Nicosia	Adults	3367
1016	FEMALE	73	89	Nicosia	Elderly	792
1017	FEMALE	74	70	Nicosia	Elderly	792
1018	FEMALE	75	88	Nicosia	Elderly	792
1019	FEMALE	45	85	Nicosia	Adults	2302
1020	MALE	33	77	Nicosia	Adults	3367
1021	FEMALE	50	65	Nicosia	Adults	2302
1022	MALE	71	85	Nicosia	Elderly	1981
1023	MALE	50	88	Nicosia	Adults	3367
1024	MALE	50	96	Nicosia	Adults	3367

Sample Counts

Count of subjects	POP_CLASS			
GENDER	Adolescents	Adults	Elderly	Grand Total
MALE	19	77	16	112
FEMALE	24	121	43	188
Grand Total	**43**	**198**	**59**	**300**

Population Real Counts

	Adolescents	Adults	Elderly	Total
MALE	52.862	259.221	31.700	343.783
FEMALE	50.254	278.554	34.061	362.869
Total	103.116	537.775	65.761	706.652

WEIGHTS (rounded)

	Adolescents	Adults	Elderly
MALE	2.782	3.367	1.981
FEMALE	2.094	2.302	792

2 http://www.cystat.gov.cy/mof/cystat/statistics.nsf/populationcondition_22main_en/populationcondition_22main_en?OpenForm&sub=2&sel=2

Subjects and Consumption template

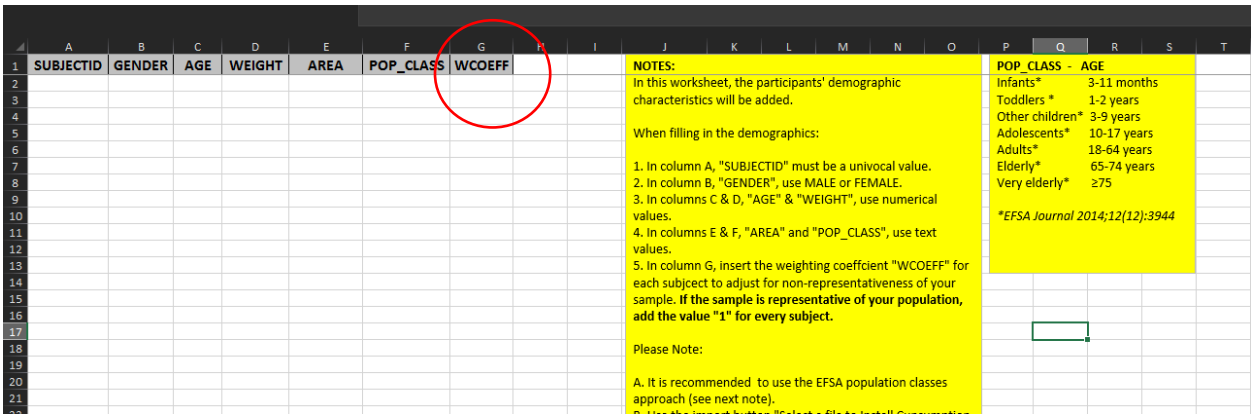
There is an updated version of the Subjects&Consumption template called *Subjects_Consumption_Template.WeightingCoefficients.xlsx* and it can be downloaded at www.improrisk.com

An additional column (column G – "WCOEFF") contains the weighting coefficient of each subject (Figure 2). Note that the values should be frequency weights[3] and not reliability weights.

An example is shown in Figure 3

IMPORTANT! When there are no weight coefficients assigned, the "WCOEFF" column is filled in with a factor of 1 for every subject (see Figure 4).

Figure 2: New template for installing new consumption data. Worksheet "Subjects"



[3] https://en.wikipedia.org/wiki/Weighted_arithmetic_mean#Frequency_weights

Figure 3: Example of weighting coefficients (see WCOEFF) in the Subjects & Consumption template- Worksheet "Subjects"

	A	B	C	D	E	F	G	H	I	J	K
1	SUBJECT	GENDE	AGE	WEIGH	AREA	POP_CLAS	WCOEFF			POP_CLASS	AGE
2	1001	FEMALE	12	81	Nicosia	Adolescents	2094			Infants	<1
3	1002	FEMALE	15	43	Nicosia	Adolescents	2094			Toddlers	1≤3
4	1003	FEMALE	62	69	Nicosia	Adults	2302			Other children	3≤10
5	1004	FEMALE	63	65	Nicosia	Adults	2302			Adolescents	10≤18
6	1005	FEMALE	64	72	Nicosia	Adults	2302			Adults	18≤65
7	1006	FEMALE	65	71	Nicosia	Adults	2302			Elderly	65≤75
8	1007	FEMALE	29	57	Nicosia	Adults	2302			Very elderly	≥75
9	1008	FEMALE	65	64	Nicosia	Adults	2302			*EFSA, 2011b	
10	1009	MALE	60	62	Nicosia	Adults	3367				
11	1010	MALE	50	104	Nicosia	Adults	3367				
12	1011	MALE	58	69	Nicosia	Adults	3367				
13	1012	MALE	45	59	Nicosia	Adults	3367				
14	1013	MALE	70	58	Nicosia	Elderly	1981				
15	1014	MALE	44	61	Nicosia	Adults	3367				
16	1015	MALE	64	91	Nicosia	Adults	3367				
17	1016	FEMALE	73	89	Nicosia	Elderly	792				
18	1017	FEMALE	74	70	Nicosia	Elderly	792				
19	1018	FEMALE	75	88	Nicosia	Elderly	792				
20	1019	FEMALE	45	85	Nicosia	Adults	2302				
21	1020	MALE	33	77	Nicosia	Adults	3367				
22	1021	FEMALE	50	65	Nicosia	Adults	2302				

Figure 4: Subjects & Consumption template. If no weigh coefficients are assigned, the "WCOEFF" column is filled in with 1

	A	B	C	D	E	F	G	H	I	J	K
1	SUBJECT	GENDE	AGE	WEIGH	AREA	POP_CLAS	WCOEFF			POP_CLASS	AGE
2	1001	FEMALE	12	81	Nicosia	Adolescents	1			Infants	<1
3	1002	FEMALE	15	43	Nicosia	Adolescents	1			Toddlers	1≤3
4	1003	FEMALE	62	69	Nicosia	Adults	1			Other children	3≤10
5	1004	FEMALE	63	65	Nicosia	Adults	1			Adolescents	10≤18
6	1005	FEMALE	64	72	Nicosia	Adults	1			Adults	18≤65
7	1006	FEMALE	65	71	Nicosia	Adults	1			Elderly	65≤75
8	1007	FEMALE	29	57	Nicosia	Adults	1			Very elderly	≥75
9	1008	FEMALE	65	64	Nicosia	Adults	1			*EFSA, 2011b	
10	1009	MALE	60	62	Nicosia	Adults	1				
11	1010	MALE	50	104	Nicosia	Adults	1				
12	1011	MALE	58	69	Nicosia	Adults	1				
13	1012	MALE	45	59	Nicosia	Adults	1				
14	1013	MALE	70	58	Nicosia	Elderly	1				
15	1014	MALE	44	61	Nicosia	Adults	1				
16	1015	MALE	64	91	Nicosia	Adults	1				
17	1016	FEMALE	73	89	Nicosia	Elderly	1				

Worksheets in ImproRisk v1.4

The following worksheet changes were made:

a) Subjects, b) Consumption, c) MeanConsumption, d) Contribution, e) Exposure- ExposureLB- ExposureMB- ExposureUB, f) GENDER- AREA- POP_CLASS, g) DrillDown.

No effect of weight coefficients is seen in the ***Comprehensive worksheet***

Worksheets *"Subjects"*

This worksheet, now additionally contains the "WCOEFF" column which corresponds to the weighting coefficient of each subject in the food survey.

Additionally, the sum of the weight coefficients is displayed, next to the Sample Size (**Error! Reference source not found.**).

Essentially, the sum of the weight coefficients will be the total population. The value *SumOfWeights* will be used for the weighted exposure statistics in the other worksheets.

NOTE!

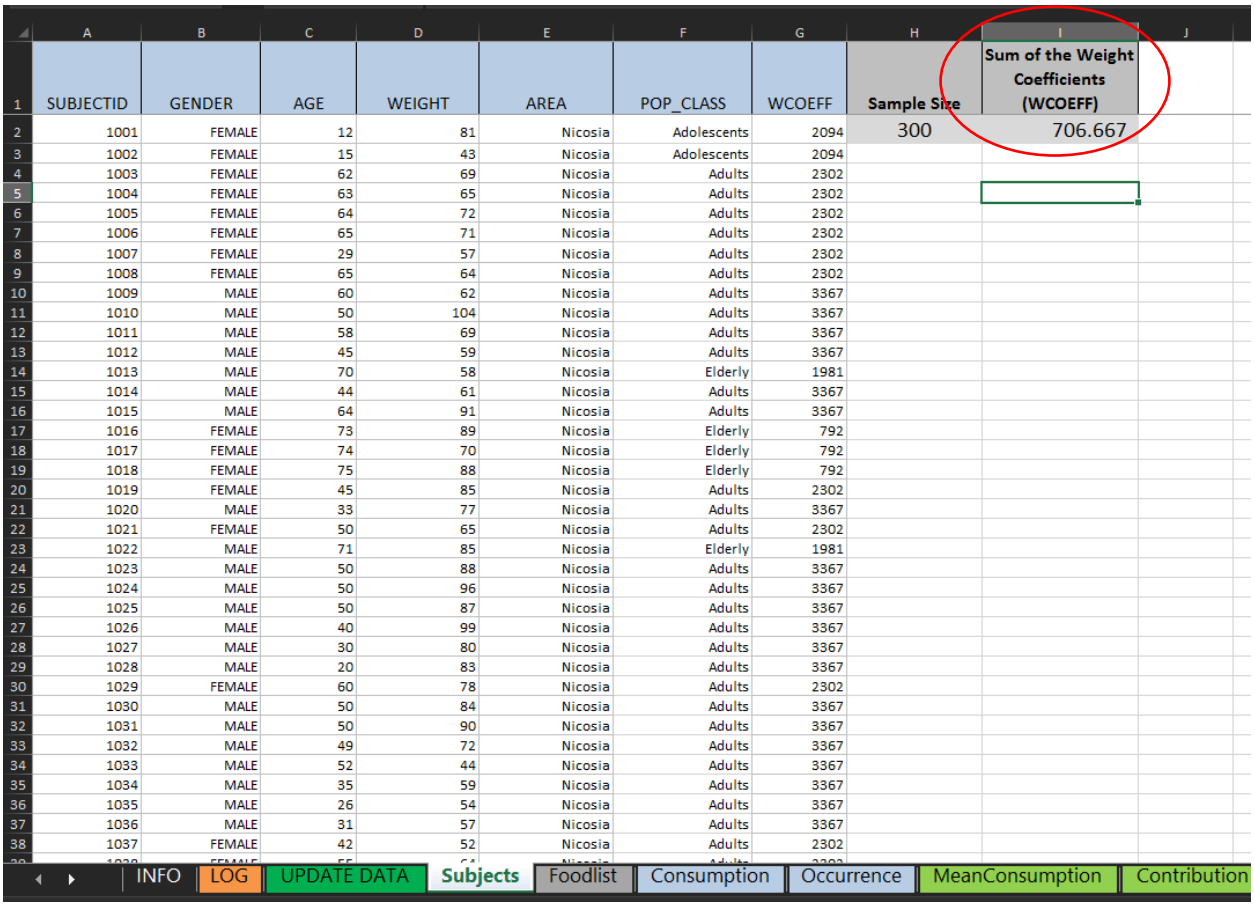
In case the user needs to "calibrate"[4] (i.e. change) the weight coefficients, it can be done through the "Subjects" worksheet. YOU SHOULD UNPROTECT the "Subjects" worksheet first (Go to Review->Unprotect. No password will be asked).

DO NOT CHANGE THE VALUES OF THE "WCOEFF" COLUMN from the Exposure worksheet; only from the "Subjects" worksheet.

When the values change in this worksheet, the model will automatically detect the changes and ask for recalculation. The model will prompt for recalculation when the user leaves the worksheet (*Figure 6*)

[4] In case of extreme values, the weight coefficient of certain subjects may be reduced to avoid inflated mean exposure estimates

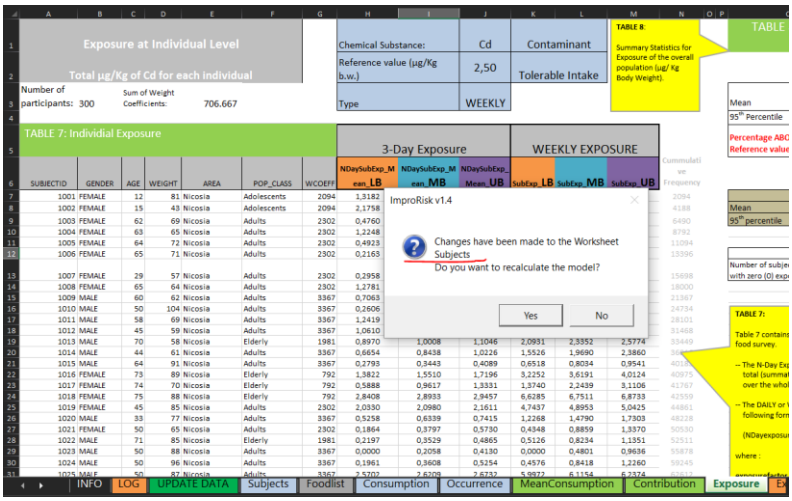
Figure 5 Worksheet "Subjects" in ImproRisk v.1.4



SUBJECTID	GENDER	AGE	WEIGHT	AREA	POP_CLASS	WCOEFF	Sample Size	Sum of the Weight Coefficients (WCOEFF)
1001	FEMALE	12	81	Nicosia	Adolescents	2094	300	706.667
1002	FEMALE	15	43	Nicosia	Adolescents	2094		
1003	FEMALE	62	69	Nicosia	Adults	2302		
1004	FEMALE	63	65	Nicosia	Adults	2302		
1005	FEMALE	64	72	Nicosia	Adults	2302		
1006	FEMALE	65	71	Nicosia	Adults	2302		
1007	FEMALE	29	57	Nicosia	Adults	2302		
1008	FEMALE	65	64	Nicosia	Adults	2302		
1009	MALE	60	62	Nicosia	Adults	3367		
1010	MALE	50	104	Nicosia	Adults	3367		
1011	MALE	58	69	Nicosia	Adults	3367		
1012	MALE	45	59	Nicosia	Adults	3367		
1013	MALE	70	58	Nicosia	Elderly	1981		
1014	MALE	44	61	Nicosia	Adults	3367		
1015	MALE	64	91	Nicosia	Adults	3367		
1016	FEMALE	73	89	Nicosia	Elderly	792		
1017	FEMALE	74	70	Nicosia	Elderly	792		
1018	FEMALE	75	88	Nicosia	Elderly	792		
1019	FEMALE	45	85	Nicosia	Adults	2302		
1020	MALE	33	77	Nicosia	Adults	3367		
1021	FEMALE	50	65	Nicosia	Adults	2302		
1022	MALE	71	85	Nicosia	Elderly	1981		
1023	MALE	50	88	Nicosia	Adults	3367		
1024	MALE	50	96	Nicosia	Adults	3367		
1025	MALE	50	87	Nicosia	Adults	3367		
1026	MALE	40	99	Nicosia	Adults	3367		
1027	MALE	30	80	Nicosia	Adults	3367		
1028	MALE	20	83	Nicosia	Adults	3367		
1029	FEMALE	60	78	Nicosia	Adults	2302		
1030	MALE	50	84	Nicosia	Adults	3367		
1031	MALE	50	90	Nicosia	Adults	3367		
1032	MALE	49	72	Nicosia	Adults	3367		
1033	MALE	52	44	Nicosia	Adults	3367		
1034	MALE	35	59	Nicosia	Adults	3367		
1035	MALE	26	54	Nicosia	Adults	3367		
1036	MALE	31	57	Nicosia	Adults	3367		
1037	FEMALE	42	52	Nicosia	Adults	2302		

INFO | LOG | UPDATE DATA | Subjects | Foodlist | Consumption | Occurrence | MeanConsumption | Contribution

Figure 6 Model detects changes in the "Subjects" worksheet and prompts for recalculation

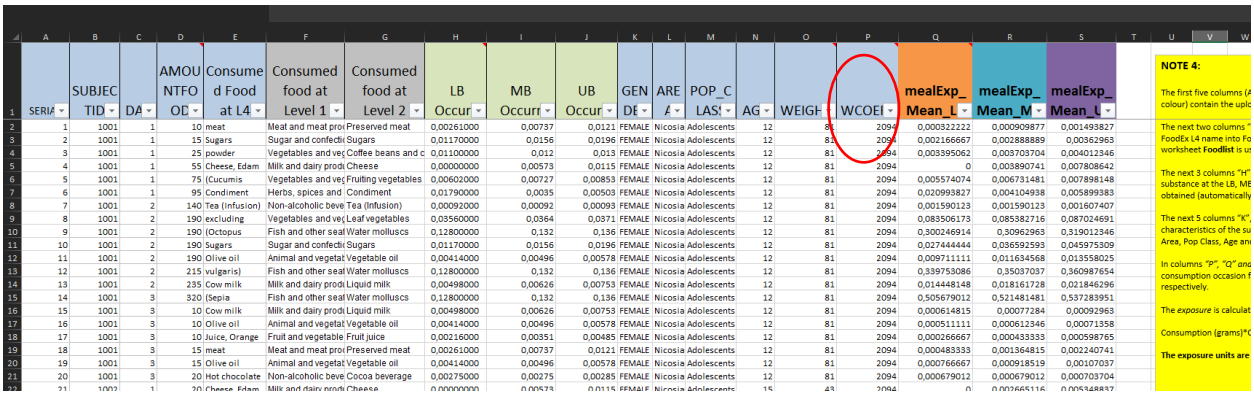


Worksheet "Consumption"

The "WCOEFF" column is also seen in the "Consumption" worksheet for each food consumption occasion (Figure 7).

Note that the "WCOEFF" column DOES NOT participate in any calculations in this worksheet. The exposure at each food consumption occasion which appears in columns *Q, R and S* is unweighted. The weighting coefficients in this worksheet are used to calculate the contribution of each food category in the worksheet "Contribution".

Figure 7: "Consumption" worksheet in ImproRisk v.1.4



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	SERIA	SUBJECTID	DA	AMOUNTFOOD	Consumed Food at L4	Consumed food at Level 1	Consumed food at Level 2	LB Occur	MB Occur	UB Occur	GENDE	ARE A	POP_CLAS	AG	WEIGH	WCOEI	mealExp_Mean_L	mealExp_Mean_M	mealExp_Mean_U		NOTE 4:		
2	1	1001	1	10	meat	Meat and meat pro	Preserved meat	0,00261000	0,00737	0,0121	FEMALE	Nicosia	Adolescents	12	81	2094	0,000322222	0,000909877	0,001493827		The first five columns (		
3	2	1001	1	15	Sugars	Sugar and confecti	Sugars	0,01170000	0,0156	0,0196	FEMALE	Nicosia	Adolescents	12	81	2094	0,002166667	0,002888889	0,00362963		colour) contain the uplo		
4	3	1001	1	25	powder	Vegetables and ve	Coffee beans and c	0,01100000	0,012	0,013	FEMALE	Nicosia	Adolescents	12	81	2094	0,003395062	0,003703704	0,004012346		The next two columns "		
5	4	1001	1	55	Cheese, Edam	Milk and dairy prod	Cheese	0,00000000	0,00573	0,0115	FEMALE	Nicosia	Adolescents	12	81	2094	0	0,003890741	0,007808642		FoodEx L4 name into Fo		
6	5	1001	1	75	(Cucumis	Vegetables and ve	Fruiting vegetables	0,00602000	0,00727	0,00853	FEMALE	Nicosia	Adolescents	12	81	2094	0,005574074	0,006731481	0,007898148		worksheet **Foodlist** is u		
7	6	1001	1	95	Condiment	Herbs, spices and	Condiment	0,01790000	0,0035	0,00503	FEMALE	Nicosia	Adolescents	12	81	2094	0,020993827	0,004104938	0,005899383		The next 3 columns "H"		
8	7	1001	2	140	Tea (infusion)	Non-alcoholic beve	Tea (infusion)	0,00092000	0,00092	0,00093	FEMALE	Nicosia	Adolescents	12	81	2094	0,001590123	0,001590123	0,001607407		substance at the LB, ME		
9	8	1001	2	190	excluding	Vegetables and ve	Leaf vegetables	0,03560000	0,0364	0,0371	FEMALE	Nicosia	Adolescents	12	81	2094	0,083506173	0,085382716	0,087024691		obtained (automatically		
10	9	1001	2	190	(Octopus	Fish and other sea	Water molluscs	0,12800000	0,132	0,136	FEMALE	Nicosia	Adolescents	12	81	2094	0,300246914	0,30962963	0,319012346		The next 5 columns "K"		
11	10	1001	2	190	Sugars	Sugar and confecti	Sugars	0,01170000	0,0156	0,0196	FEMALE	Nicosia	Adolescents	12	81	2094	0,027444444	0,036592593	0,045975309		characteristics of the su		
12	11	1001	2	190	Olive oil	Animal and vegetat	Vegetable oil	0,00414000	0,00496	0,00578	FEMALE	Nicosia	Adolescents	12	81	2094	0,009711111	0,011634568	0,013558025		Area, Pop Class, Age an		
13	12	1001	2	215	vulgaris)	Fish and other sea	Water molluscs	0,12800000	0,132	0,136	FEMALE	Nicosia	Adolescents	12	81	2094	0,339753086	0,35037037	0,360987654		In columns "P", "Q" and		
14	13	1001	2	235	Cow milk	Milk and dairy prod	Liquid milk	0,00498000	0,00626	0,00753	FEMALE	Nicosia	Adolescents	12	81	2094	0,014448148	0,018161728	0,021846296		consumption occasion f		
15	14	1001	3	320	(Sepia	Fish and other sea	Water molluscs	0,12800000	0,132	0,136	FEMALE	Nicosia	Adolescents	12	81	2094	0,505679012	0,521481481	0,537283951		respectively.		
16	15	1001	3	10	Cow milk	Milk and dairy prod	Liquid milk	0,00498000	0,00626	0,00753	FEMALE	Nicosia	Adolescents	12	81	2094	0,000614815	0,00077284	0,00092963		The *exposure* is calculat		
17	16	1001	3	10	Olive oil	Animal and vegetat	Vegetable oil	0,00414000	0,00496	0,00578	FEMALE	Nicosia	Adolescents	12	81	2094	0,000511111	0,000612346	0,00071358				
18	17	1001	3	10	Juice, Orange	Fruit and vegetable	Fruit juice	0,00216000	0,00351	0,00485	FEMALE	Nicosia	Adolescents	12	81	2094	0,000266667	0,000433333	0,000598765		Consumption (grams)*C		
19	18	1001	3	15	meat	Meat and meat pro	Preserved meat	0,00261000	0,00737	0,0121	FEMALE	Nicosia	Adolescents	12	81	2094	0,000483333	0,001364815	0,002240741				
20	19	1001	3	15	Olive oil	Animal and vegetat	Vegetable oil	0,00414000	0,00496	0,00578	FEMALE	Nicosia	Adolescents	12	81	2094	0,000766667	0,000918519	0,00107037		**The exposure units are**		
21	20	1001	3	20	Hot chocolate	Non-alcoholic beve	Cocoa beverage	0,00275000	0,00275	0,00285	FEMALE	Nicosia	Adolescents	12	81	2094	0,000679012	0,000679012	0,000703704				
22	21	1002	1	20	Cheese, Edam	Milk and dairy prod	Cheese	0,00000000	0,00573	0,0115	FEMALE	Nicosia	Adolescents	15	43	2094	0	0,002665116	0,005348837				

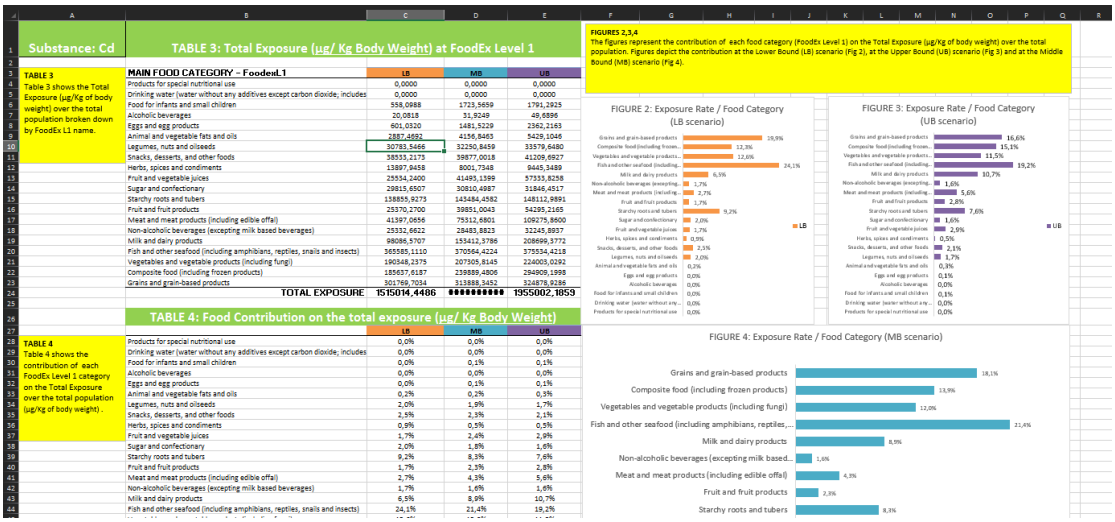
Worksheet "Contribution"

The look and feel of the Contribution worksheet has not changed, as shown in Figure 8. Nevertheless, the calculations performed have changed to consider the weighting coefficient of each subject. In Table 3, it is noticeable that the numbers have been increased considerably since the "WCOEFF" values are included in the formulas. The total exposure calculated here, is the estimated total exposure for the entire population. The same applies for Table 6.

Table 4 still presents the contribution of each FoodEx Level1 food category to the total exposure.

Table 5 now depicts the estimated mean exposure per day per person in the entire population. The formulas have been updated to *TotalExposure/SurveyDays/SumoftheWCOEFF*, instead of the *TotalExposure/SurveyDays/SampleSize.*

Figure 8: "Contribution" worksheet in ImproRisk v.1.4



Worksheet "MeanConsumption"

The mean (average) daily consumption is calculated in this worksheet for each food at FoodEx Level 1 (20 categories), as shown in Figure 9.

Table 1

In Table 1, there were no alterations in the calculations. The table still calculates the mean consumption, throughout the food survey period, for each individual in the sample and for each category at FoodEx Level 1.

Table 2

A. An additional column "Number of Consumers in the Population" is added in Table 2, which shows the *projected* number of consumers in the population. This number is simply the summation of the weight coefficients of the sample participants that consumed each food at Level 1.

The formula used is $\sum_{i=1}^{k} w_i$,

where k is the number of subjects who consumed the food, and w_i is the respective weight coefficient of the i_{th} consumer.

B. The *"Consumer based (gr)"* calculations (next column) have changed and

the formula used is $\frac{\sum_{i=1}^{k} y_i w_i}{\sum_{i=1}^{k} w_i}$,

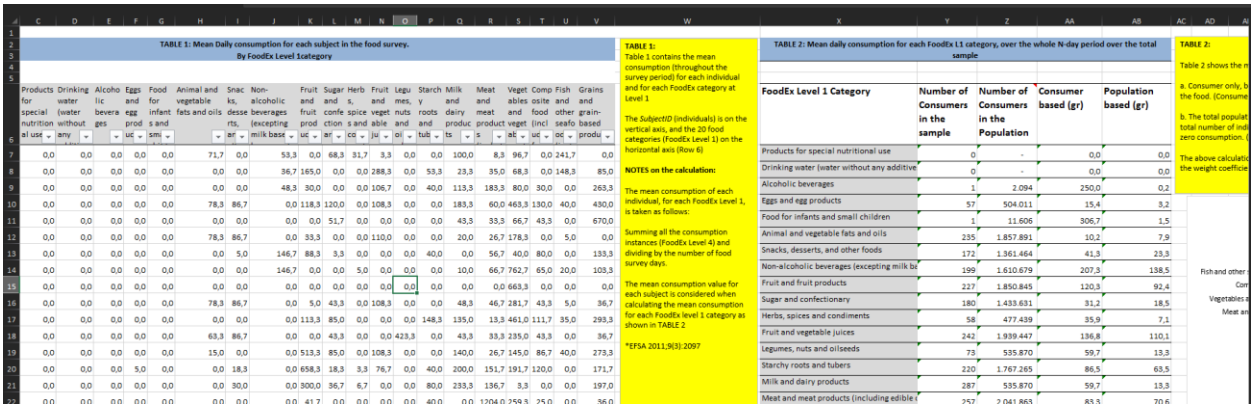
where k is the number of subjects who consumed the food, y_i is the mean consumption of the i_{th} consumer, and w_i is the respective weight coefficient of the i_{th} consumer. Essentially, the formula calculates the total weighted (i.e. multiplied by the weight coefficient) mean daily food consumption for each individual, and divides by the number of consumers in the population.

C. The *"Population Based (gr)"* divides over the whole population (i.e. Sum Of Weights)

the formula used is $\frac{\sum_{i=1}^{k} y_i w_i}{\sum_{i=1}^{n} w_{i,}}$,

where k is the number of subjects that consumed the food, y_i is the mean consumption of the i_{th} consumer, w_i is the respective weight coefficient of the i_{th} consumer, and n is the total number of participants in the food survey.

Figure 9: Example of the "MeanConsumption" worksheet in the ImproRisk v.1.4



Worksheet "*Exposure*"

The calculations in Table 7 have not changed. The Weekly or Daily exposure is calculated for each individual in the sample. The WCOEFF column (new column added) is not considered in the calculations of table 7 (Figure 10).

However, the WCOEFF column is used to calculate the weighted mean exposure in Table 8.

A. The weighted mean exposure

The weighted mean exposure is calculated using the following formula:

Weighted mean exposure = $\frac{\sum_{i=1}^{n} y_i w_i}{\sum_{i=1}^{n} w_{i,}}$

where n is the number of subjects in the sample, y_i is the exposure for the i_{th} subject, and w_i is the respective weight coefficient for the i_{th} subject.

Essentially, instead of each data point (exposure of each individual in the sample) contributing equally to the final average, each individual contributes differently based on the weight coefficient. If all the weights are equal (e.g. 1), the weighted mean is the same as the unweighted mean.

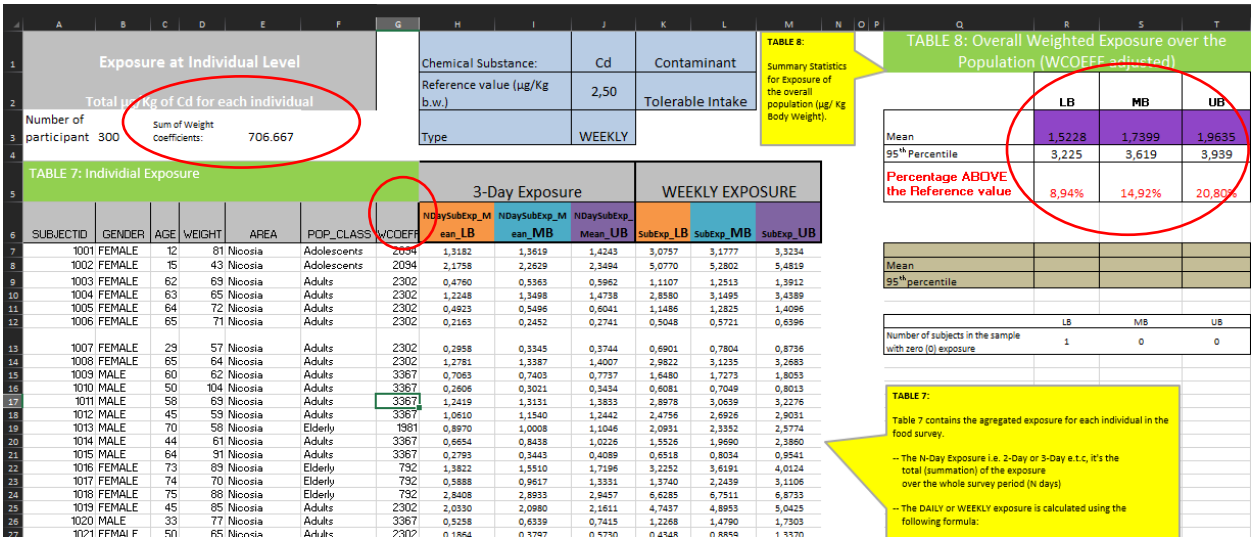
B. The 95th percentile

Due to the complexity of the calculations and formulae used when weighting factors are involved, the 95th percentile of exposure is calculated internally using a VBA script when either new Consumption or Occurrence data are installed.

C. Percentage above the reference value

Similarly, a percentage of the total population that exceeds the reference value is calculated and the weighting factor is also considered.

Figure 10: Worksheet "Exposure" in ImproRisk v.1.4



The summary statistics in Table 13 of the worksheet are all weighted (Figure 11).

The frequency weights (i.e."repeat"-type weights - integers counting the number of occurrences for each observation) and the estimators used are considered unbiased[5]. Analogous calculations are performed in the worksheets "ExposureLB" and ExposureUB".

The 25th percentile (Q1), the median (Q2) and the 75th percentile (Q3) are calculated internally using a VBA script. The IQR (Inter Quartile Range) is Q3-Q1, which is used in the calculation of the bins width for the histogram in Figure 8 in the worksheet.

The weighted standard deviation is calculated using the following procedure:

The weighted variance[6] is

$$s^2 = \frac{\sum_{i=1}^{n} w_i * (y_i - \mu)^2}{\sum_{i=1}^{n} w_i}$$

or…

$$s^2 = \frac{\sum_{i=1}^{n} (w_i * y_i^2) - \mu^2 * \sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} w_i - 1}$$

where n is the number of subjects in the sample, y_i is the exposure for the i_{th} subject, w_i is the respective weight coefficient for the i_{th} subject, and μ is the weighted mean exposure.

therefore, the weighted standard deviation is

$$s = \sqrt{\frac{\sum_{i=1}^{n} (w_i * y_i^2) - \mu^2 * \sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} w_i - 1}}$$

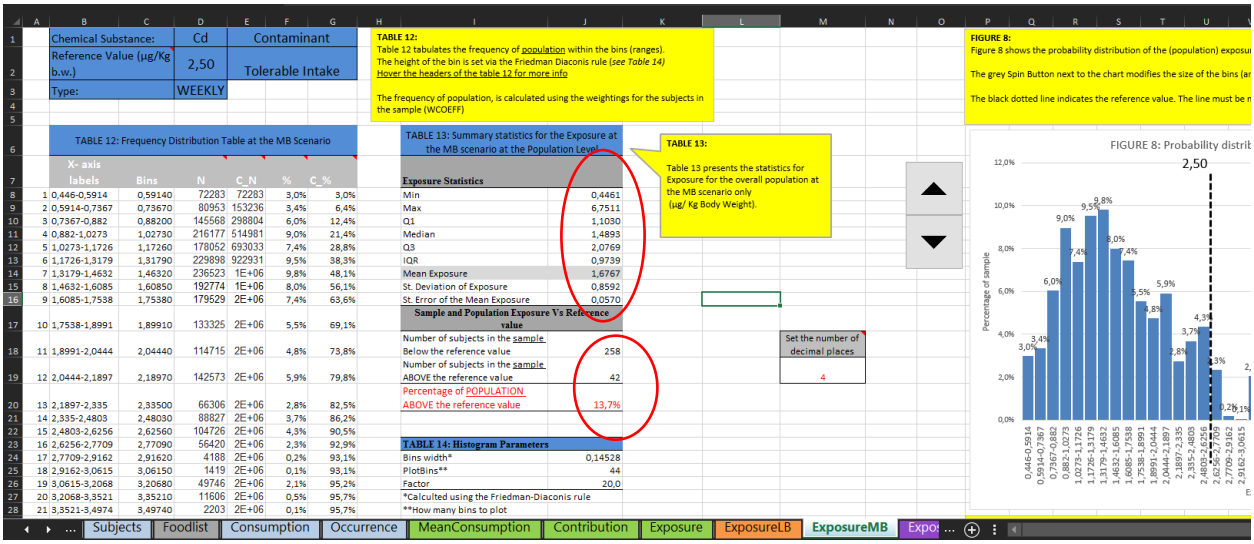
The standard error[7] of the weighted mean exposure is given by

$$se = \frac{s\sqrt{\sum_{i=1}^{n} w_i}}{\sum_{i=1}^{n} w_i}$$

where s is the weighted standard deviation, n is the number of subjects in the sample, and w_i is the weight coefficient for the i_{th} subject.

[5] https://en.wikipedia.org/wiki/Weighted_arithmetic_mean#Frequency_weights
[6] https://en.wikipedia.org/wiki/Weighted_arithmetic_mean#Frequency_weights
https://stats.stackexchange.com/questions/47325/bias-correction-in-weighted-variance
[7] http://www.analyticalgroup.com/download/WEIGHTED_MEAN.pdf

Figure 11 Exposure MB worksheet in ImproRisk v.1.4



Moreover, Table 13 shows the number of subjects in the sample that are below or above the reference value and additionally the weighted proportion of the population above the reference value.

Worksheets "GENDER", "AREA" and "POP_CLASS"

Analogous calculations for the weighted mean exposure, weighted standard deviation and standard error of the exposure, are calculated across each demographic.

Note, that the estimates in these worksheets are based on the Middle Bound (MB) scenario of exposure.

Worksheet "DrillDown"

The use of *calculated fields* functionality in excel's *pivot tables* lead to the creation of a file under the name "ExposureMB_weighted", which contains the weighted exposure of the subjects (Figure 12). Mean exposure values shown on the pivot table and pivot chart are weighted.

The first visualisation shown in the previous version of ImproRisk (v1.3.5) for the exposure across gender and LB-MB-UB scenarios, is removed from this version.

Figure 12: "DrillDown" worksheet in ImproRisk v.1.4