Instructions for export of analysis results from the ESCAPE-NET server

May 11, 2020

Due to the secure cloud environment users of the ESCAPE-NET server are not able to export any analysis results, metadata information or figures from the server without authorization from the WP2 team and the help of the consortium data manager.

For a description on how to request data export, please refer to the standard operating procedure for user and data management in BC|GENOME.

1 Rules for export of result files

As a general rule, only analysis results, aggregated tables or figures for which it is not possible to identify individual units can be exported from the ESCAPE-NET server. This is true even when the person identification numbers have been pseudonymized and also even if the tokens (i.e. the tokens generated in the pseudonymization) have been removed.

> **The results exported, should in principle be in a form such that they can be used in a publication.**
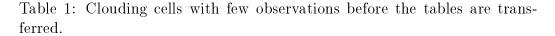
For tables, the rule of thumb is that tables should contain at least 3 observations, but the level of aggregation must always depend on a concrete assessment of what the table contains. If you have any doubts whether 3 observations is enough to ensure anonymity, then further aggregation should be done.

> **Tables must contain at least 3 observations per cell.**

If fewer observations are contained in the tables, a further aggregation of these tables is needed. Alternatively, you can cloud cells with few observations before the tables are transferred. Clouding can e.g. be conducted by deleting the content of the cell or by replacing it with a marking of, e.g. "<3" as in table 1.

Duration of smoking in pregnancy	Amount of smoking			
	Mild	Moderate	Heavy	Total
-18 weeks	15	6	2	23
18-31 weeks	2	1	11	14
32+ weeks	5	9	15	29

Duration of smoking in pregnancy	Amount of smoking			
	Mild	Moderate	Heavy	Total
-18 weeks	15	6	<3	23
18-31 weeks	<3	<3	11	14
32+ weeks	5	9	15	29

Table 1: Clouding cells with few observations before the tables are transferred.

Please be aware that it does not suffice to cloud a single cell with too few observations if the original value can be backtracked by the marginal totals (as is the case in first row in above figure, marked with a red circle).

Alternatively, categories with too few observations can be combined such as in table 2, where the categories "None" and "1-3 years" have been combined into "<3 years".

Values describing distributions, e.g. max and min and percentiles such as the median may only be returned if there is no risk of identifying individuals.

One should pay particular attention to whether results contains outliers, e.i. extreme values (e.g. min and max) and possibly remove them before sending. If max and min values are to be exported, observations at both ends should be close to each other so that no observations stands out.
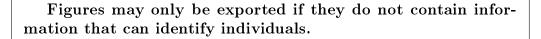
Highest attained educational level	Individual level disposable household income in quartiles				Total
	First	Second	Third	Fourth	Total
None	1	1	8	18	28
1-3 years	2	6	14	17	39
>3 years	8	18	16	34	46
Academic education	13	9	7	7	34

↓

Highest attained educational level	Individual level disposable household income in quartiles				Total
	First	Second	Third	Fourth	Total
<3 years	3	7	22	35	57
>3 years	8	18	16	4	46
Academic education	13	9	7	7	34

Table 2: Combining cells with few observations before the tables are transferred.

> **Max, min and percentiles, such as the median may only be exported if there is no risk of identification of individuals.**

Figures may only be returned if they do not contain identifiable information. One must also pay special attention to outliers / extreme values. In addition, it must be ensured that the figures do not contain embedded metadata. The latter can be ensured by exporting the figures in pdf format.

> **Figures may only be exported if they do not contain information that can identify individuals.**

Programs (code) and log files can be exported but one should be very careful to confirm that they do not contain data on individual level. E.g.:

- Code containing data on individual level, e.g. written into a condition

- Logs that list observations. Some procedures list e.g. a small sample of the data set used for the analysis

3

Generally, the best practice is to keep program on the server and carry out all analysis-work here.

Final advice's for the user of the ESCAPE-NET server

- Always thoroughly check all the files that you want to export. If in doubt, discuss the case with your colleges or contact data manager (simon.mathis.koenig@regionh.dk). Do not share identifiable information from the server in emails. If you want to discuss a concrete example (e.g. a table), you should instead use the sharing options in the BC|GENOME.

- Restrict the export to what is really needed

    - Is it necessary, for example, to export programs and logs? - all analysis should anyway be carried out on the server.

    - Is it necessary to export preliminary results?