DEEP SEAS COMMON DATA BASE GUIDELINES[1]

I. GENERAL COMMENTS

1. Names of variables

Each input variable name corresponds with relevant number of the question, it consists of abbreviation of the questionnaire section name, lower dash and number or character of the question, e.g., the question: "WB_1. How is your health in general?" – variable name: WB_1 or the question "HD_2. Who are these people? – HD_2a_1. Household member" – variable name: HD_2a_1.

2. Variable types

There are two types of variables:

1) Input variables – data transferred directly from the questionnaire.

2) Target variables – data used in analyses created on the basis of input variables after various transformations.

Sometimes target variables serve as input variables to construct more complex target variables.

Data base consists of 147 input variables and 44 target variables.

3. Global codes

Code	Responses	Comments
0	Does not apply	Results only from skips or as natural 0 in scale variables
sysmis (system-missing)	No data	No valid data or invalid data In target variables sysmis may mean missing values resulting from variables transformations and/or extremely high, implausible information on volumes consumed.

4. Checking data consistency

In case of the majority of related variables consistency was checked and inconsistent answers were corrected if it was possible and feasible. For instance, for respondents who declare generic frequency of drinking (F) lower than frequency of drinking of the most frequently drunk alcohol beverage (BSQF) the generic frequency was replaced by frequency of drinking of this beverage. In some cases, inconsistency was tolerated for example between BSQF and RSOD questions.

5. Data weighting

In all samples weights were constructed to adjust to population data based on three dimensions, separate: sex, three age categories and region.

II. COMMENTS TO PARTICULAR SECTIONS

1. Establishing drinking/abstainer status (F)

Input variable:

Variable name	Variable label
F_1b	Frequency of drinking

Target variable:

Variable name	Variable label
Consumers	alcohol consumers

- Abstainers – all who confirmed either 10 or 11 among response categories provided.
- Drinkers – all who confirmed either of 1 to 9 in responses to F1, no matter if any drinking was reported in BSQF section.

2. Generic frequency of drinking (F)

Input variable:

Variable name	Variable label
F_1b	generic frequency of drinking

Target variable:

Variable name	Variable label
rf_1b	generic frequency of drinking (number of days in year)

The question on overall or generic frequency of drinking was intended to capture drinking of any alcoholic beverage; that means beer, wine, spirits, or any other alcoholic beverage even in small volumes. This question was asked after the BSQF section to avoid a filter effect which could increase the estimated abstinence rate (Sierosławski, Foster, and Moskalewicz 2013).

Frequency of drinking scale was converted into number of drinking days in the following way:

Frequency of drinking scale	Range	Middle point (number of drinking days)
1) Every day	365	365.0
2) 5 – 6 times a week	364 – 207	285.5
3) 3 - 4 times a week	156 – 208	182.0
4) 1 - 2 times a week	52 – 104	78.0
5) 2 - 3 times a month	24 – 36	30.0
6) Once a month	12	12.0
7) 6 - 11 times a year	6 – 11	8.5
8) 2 - 5 times a year	2 – 5	3.5
9) Once a year	1	1.0
10) No during last 12 moths	0	0.0
11) Never during in lifetime	0	0.0

3. **Beverage Specific Quantity-Frequency (BSQF)**

To reduce the confusion related to the concept of "average" the adjective "usual" volume consumed was applied. In most countries, with the exception of Denmark and Norway, the question consisted of two steps. First respondents selected a container i.e. reporting units (e.g. 33 cl cans/bottles or 50cl cans/bottles), which had been predefined specifically by country, based on the most common glasses or bottles or cans used in that country. Then they reported number of units they usually consume on one drinking day.

The usual intake of particular alcohol beverage was calculated in terms of centilitres of 100% alcohol. The capping of extreme volumes was applied at the level of 50 cl of 100% alcohol per drinking day: all values between 50 and 80 cl were set to 50 cl of 100% alcohol while values above 80 cl were set to missing.

Input variables:

Frequency of drinking:

Variable name	Variable label
bsqf_1b	frequency of beer drinking
bsqf_3b	frequency of wine drinking
bsqf_5b	frequency of spirits drinking
bsqf_7b	frequency of drinking of additional beverage

Target variables:

Frequency of drinking in days:

rbsqf_1b	frequency of beer drinking (number of days in year)
rbsqf_3b	frequency of wine drinking (number of days in year)
rbsqf_5b	frequency of spirits drinking (number of days in year)
rbsqf_7b	frequency of drinking of additional beverage (number of days in year)

Quantity of drinking in terms of 100% alcohol:

Variable name	Variable label
bsqf_2a	quantity of beer on usual drinking day (cl of 100% alc)
bsqf_4a	quantity of wine on usual drinking day (cl of 100% alc)
bsqf_6a	quantity of spirits on usual drinking day (cl of 100% alc)
bsqf_8a	quantity of additional beverage on usual drinking day (cl of 100% alc)
cbsqf_2	quantity of beer on usual drinking day (cl of 100% alc) – capped
cbsqf_4	quantity of wine on usual drinking day (cl of 100% alc) – capped
cbsqf_6	quantity of spirits on usual drinking day (cl of 100% alc) – capped
cbsqf_8	quantity of additional beverage on usual drinking day (cl of 100% alc) – capped

Country specific variables including type of container and number of containers are not included into this data base. Instead, there are target variables cbsqf_2; cbsqf_4; cbsqf_6; cbsqf_8; containing usual quantities of three basic and one additional alcohol beverages consumed on drinking day recalculated into cl of 100% alcohol.

Target variables:

Variable name	Variable label
bsqf_beer	beer annual consumption in cl of 100% alc
bsqf_wine	wine annual consumption in cl of 100% alc
bsqf_spir	spirits annual consumption in cl of 100% alc
bsqf_adbev	additional beverages annual consumption in cl of 100% alc
bsqf_alc	alcohol annual consumption in cl of 100% alc
bsqf_alc1	alcohol annual consumption in cl of 100% alc - after capping (18250 cl. 100% alc)
cbsqf_beer	beer annual consumption in cl of 100% alc – capped
cbsqf_wine	wine annual consumption in cl of 100% alc – capped
cbsqf_spir	spirits annual consumption in cl of 100% alc – capped
cbsqf_adbev	additional beverages annual consumption in cl of 100% alc - capped
cbsqf_alc	alcohol annual consumption in cl of 100% alc – capped
c1cbsqf_alc	alcohol annual consumption in cl of 100% alc with firts capping - after second capping (18250 cl. 100% alc)

The target variables presented above constitute estimation of annual consumption of beer, wine and spirits (and additional beverage where relevant) and of overall annual consumption based on uncapped volumes (first five variables) as well as annuals volumes after capping with the extreme annual volumes of overall alcohol consumption (above 182.5 litres of 100% alcohol) capped again.

The algorithm of calculation of core target variables related to annual alcohol consumption was as follows:

The frequencies of drinking of each particular beverage were converted into number of drinking days in the same way as frequency of drinking in generic frequency of drinking section (see point 2).

To derive the beverage-specific usual quantity of alcohol drunk in centilitres of 100% alcohol, the number of units of each alcoholic beverage the respondent reported was first multiplied by a predefined volume of a unit (e.g. 33 cl or 50 cl; separately for each beverage). The result of this calculation was then multiplied by alcohol content (in per cent) of the beverage (separately for each beverage). To derive beverage-specific annual consumption, in cl of 100% alcohol, the number of drinking days of a given beverage was multiplied by the usual quantity in cl of 100% alcohol. To estimate total average annual alcohol consumption across beverages, the annual volumes of consumption of each alcoholic beverage in centilitres of 100% alcohol were summed. If data on consumption of a particular alcoholic beverage were missing, the estimation was done on the basis of the available data. That means the missing data were set to zero to prevent calculating missing data on overall alcohol consumption. In the case of respondents who were not able to provide information on drinking of any alcoholic beverage or provided implausible volumes above 80 cl of pure alcohol and were not abstainers, the data on alcohol consumption was considered missing.

4. **Risky Single Occasion Drinking (RSOD)**

Risky Single Occasion Drinking (RSOD) is defined as consumption of at least 60 grams of 100% alcohol by males on one occasion and at least 40 grams of 100% alcohol by females. These volumes for an average human being may cause intoxication expressed as 0.5‰ BAC. The threshold of 60/40 grams of pure alcohol on one occasion seems to be too low to identify episodes of higher intoxication e.g. BAC over 1‰.To explore the more extreme end of the distribution of volume consumed on one occasion, a question about frequency of drinking 120+ grams of pure alcohol for males and 80+ grams of pure alcohol for females was also applied.

The frequency of RSOD was investigated using a standard frequency scale used throughout all questions on frequency of alcohol consumption.

Frequency of drinking scales were converted into number of drinking days in the same way as frequency of drinking in generic frequency of drinking section (see point 2).

Additionally, the question on maximum number of standard drinks on one occasion in the last 12 months was applied to get a reference point for the calculation of volume of alcohol attributed to RSOD. The data base includes that volume calculated in grams of 100% alcohol (cRSOD_1a). The capping procedure was used, and values higher than 400 were set to 400 grams of pure alcohol The volume in grams was also converted into the volume in cl (rsod_1cl) i.e., to 50 cl.

Input variable:

Variable name	Variable label
rsod_2b	Frequency of drinking 4/6+ drinks
rsod_7b	Frequency of drinking 8/12+ drinks
RSOD_1a	Max vol of 100% alc in grams

Target variable:

Variable name	Variable label
cRSOD_1a	Max vol of 100% alc in grams – capped
rrsod_2b	frequency of drinking of 40/60 g alc (number of days in year)
rrsod_7b	frequency of drinking of 80/120 g alc (number of days in year)

5. **Unrecorded alcohol (UA)**

To estimate share of unrecorded alcohol in overall consumption relevant variables reported in terms of litres of beverages were converted to 100% alcohol.

6. **Gender (SD_1_gender)**

Question on gender included category "Other". As the RSOD question was gender-specific, therefore the respondents who select category "other" were asked to make a choice between male or female version of RSOD question. The new gender variable was created by combining gender selected in SD_1 and choice made before RSOD section.

Input variable:

Variable name	Variable label
SD_1	What is your gender?
RSOD_1_Other	The following questions consider sex-specific standards of alcohol consumption. Please let us know if you wish to respond to the male or female version?

Target variable:

Variable name	Variable label
SD_1_gender	gender as filter for RSOD

7. Social class (SD_10 – SD_18)

Social class of respondents is based on their current occupation. In case of respondents who were not professionally active their social class was based on their most recent employment, and for those who had never held any paid job, their social class was based on employment of the head of their household.

Input variable:

Variable name	Variable label
sd_10	professionally active vs non-active
sd_11	Current occupation status
sd_12	Respondent self-employed now
sd_13	Professionally non-active
sd_14	Paid work in the past
sd_15	Occupation status of non-working
sd_16	Respondent self-employed in the past
sd_17	Occupation status of a spouse/parent
sd_18	Is your spouse/parent self-employed or employed

Target variable:

Variable name	Variable label
social_class	social class

8. Household income (SD_20, SD_20a)

Household net income was investigated using two consecutive questions. Firstly, respondents were asked directly about exact household net income expressed in national currency. For those who were not able to provide exact figures the subsequent question provided opportunity to respond selecting on the scale corresponding income category. The country-specific income categories were converted into figures taking a midpoint as approximation of income. Midpoint for the lowest open category was equal three quarters of that category and for the highest open category a midpoint was 50% higher. The same thresholds were applied to the variable containing exact household income i.e. all incomes below 75% of the midpoint for the lowest category were up-graded to that level and all incomes higher than 50% of the highest category were capped to that level. Finally, both income variables were integrated in a variable sd_20C. In those countries that reported annual income identified on the basis of variable SD_20_income_type2 relevant income figures were divided by 12 (sd_20month). Household income per household member was calculated dividing reported income by a number of household members.

Source variable:

Variable name	Variable label
sd_20	Household income in national currency
sd_20A	Household income in national currency in ranges
SD_20_income_type	Yearly – monthly

Target variables:

Variable name	Variable label
sd_20B	Household income calculated on the basis of ranges
sd_20C	Integrated household income (SD_20 and SD_20B)
sd_20month	Monthly household income final
inc_mth	Monthly income per capita of household members final

In the DEEP SEAS survey carried out in 2021 the Slovenian income had to be set to missing as the Slovenian questionnaire did not specify whether the income had to be reported in terms of annual or monthly income.

In Iceland in the same DEEP SEAS survey, the first question on precise income (sd_20) was on annual income while the next one on income in ranges (sd_20a) was on monthly income. Therefore, sd_20C, which mixes up annual and monthly incomes should be ignored for Iceland.