WHO EUROPEAN CHILDHOOD OBESITY SURVEILLANCE INITIATIVE (COSI)

GUIDELINES ON DATA PROCESSING AND CLEANING

June 2022

This document illustrates the procedures for data processing and cleaning that should be carried out at national level as part of the data validation process. This document covers the following functions:

1. Check data for completeness, plausibility and consistency;
2. Match the school, child and family datasets;
3. Assess the nutritional status of children using the international growth reference curves (WHO and IOTF);
4. Calculate indicators to assess data quality: the rate of participation in the study, the completeness of key information on children and the level to which the children dataset matches family and school datasets;
5. Create the datasets requested by WHO Regional Office.

LibreClinica (LC), the system provided by WHO to COSI participating countries to computerize data, automatically performs standard controls on plausibility and consistency of entered data and linkage of children's, family's and school datasets. Countries that use LC will carry out part of the validating process within the LC environment. Nevertheless, it is highly recommended to carry out a comprehensive check for completeness, plausibility and consistency of data once they are extracted from LibreClinica, such as verifying datasets matching.

The procedures illustrated in points 1-4 will be carried out at European level once validated national data files will be sent to WHO Regional Office in order to increase the comparability among countries and over time.

Table of Contents

1. Check data for completeness, plausibility and consistency

1.1 Main steps

1) Check that the number of observations in the datasets is as expected and that each observation is univocally identified in the dataset by a code (i.e. all paper forms were properly coded and digitalized). If two records have the same code, ascertain the reason and assign to each record a unique code. The reason can be: i) a duplication of the same record; ii) an error occurred when data was digitalized (data entry error) or iii) an error occurred when the code was recorded on the paper form (data recording error). In the latter case, assign a new code to one of the two records in the dataset and register the code on the corresponding paper form.

2) Check that variables format, values and labels are as expected on the base of the questionnaire and on the data codebook used to digitalize the information collected through paper forms.

Analyse the frequency distribution of the variables by calculating frequency tables of categorical variables and measures of central tendency (mean, median, etc.) and measures of spread/dispersion (variance, interquartile range) of interval variables. If a distribution seems implausible (based on previous rounds findings and/or other reliable data sources), carry out a more in-depth analysis to understand what might have happened.

In the case of interval variables (such as children anthropometric measures, parents' body weight and height, etc.), a unique definition for extreme/implausible values cannot be provided. Each country will individually determine values that need to be treated as potential extreme/implausible values. The suggestion is to carry out an in-depth analysis of the distribution, using also some graphical techniques such as box plot, normal plot, etc.

In case of out of range or potential extreme/implausible values, handle them as described in point 4).

Create a new variable at the end of the dataset (called "Notes") where you will annotate any specific problem that may arise from data checking.

3) Perform procedures to check for out of range or extreme/implausible values of some specific variables as described in paragraph 1.2. In case of out of range or extreme/implausible values, handle them as described in point 4).

4) Out of range or extreme/implausible values could be the result of an error either in recording the data on the paper form or on the electronic data entry system, or could be associated with a value that, despite being extreme, is true. Data checking aims at understanding if an error

occurred and if it is possible to correct it. To this extent, handle out of range or extreme/implausible values as follows:

- ✓ Double-check the value on paper form and/or with the examiner who collected the data to understand if a data entry or a recording error occurred;

- ✓ If an error occurred and the correct value can be determined, change it accordingly;

- ✓ If it is not possible to determine if an error occurred or to identify the correct value, report the problem on the "Notes" variable.

5) Check the level of missing data for each variable and assess if a more in-depth analysis is needed (such as double-checking the paper version of the questionnaire or ask for explanations to the examiners).

Pay particular attention to identifying and checking observations that have a high number of blank variables or that have missing data on key-variables (such as child's anthropometric measurements, child's and parents' socio-demographic characteristics) and report the problem on the "Notes" variable.

As regards the child's record form dataset,

- ✓ Please pay particular attention to blank fields on the following variables: child's sex, child's date of birth or child´s age in months, date of measurement, child's body weight, child's height, and clothes worn when the child was measured. All these data are requested in order to assess child's nutritional status. For all children who agreed to be measured, in case of missing data on one or more of these variables, the paper questionnaire has to be checked and data has to be recovered if possible.

6) Carry out procedures to check consistency between variables as described in paragraph 1.3. In case of inconsistencies, handle them as described in point 7).

7) Handle inconsistencies among variables:

- a. Double-check incoherent values on paper form in order to see if a data entry error occurred. If this is the case, correct the values accordingly. If not, go at point b.

- b. Ask the examiner who took the measurements if she/he can provide an explanation. If it the inconsistency cannot be solved, go at point c.

- c. Assess if it is possible to establish which variable value is more reliable or reasonable (based on experience gained during previous data collections or other similar studies) and correct data accordingly. If this is not possible, go at point d.

- d. If the inconsistency cannot be solved, report the inconsistency on the "Notes" variable.

e. Similar inconsistencies should be handled using the same approach, also reporting in detail how each type of inconsistency has been solved (use "Notes on COSI 2021-23 Datasets" template).

1.2 Additional information on the procedures to check value plausibility of variables

Child's record form

A. Child's identification code (child's ID)

Check that the child's ID was defined according to WHO 2021/23 data codebook - 13 digits as follows: country code (3 digits), the year of the data collection (2 digits), school identification code (4 digits), class grade (1 digit), class identification code (1 digit) and child's code (2 digits). See COSI Manual of Data Collection Procedures for more details.

If a different definition was used, make a note to WHO Regional Office

B. (M2) Child's date of birth

✓ Check that the year of birth is compatible with the expected age of children. A common error that may occur is to register the date when the paper form was filled instead of the child's date of birth.

✓ Check if the dataset contains an excessive number of children born on the same day or on the same month. If this is the case, make a more in-depth analysis in order to find a reasonable explanation (i.e. understanding if a data entry or recording error occurred).

✓ After verifying the date of measurement, calculate the age in years or months and check for extreme/implausible values. The age can be calculate using the following formulas:
- Age in years = (date of measurement-date of birth)/365.25
- Age in months = (date of measurement-date of birth)/ 30.4375

✓ Based on children age in years or months, calculate age groups (6.0-6.9, 7.0-7.9, 8.0-8.9, 9.0-9.9 years old). Check that the number of children belonging to targeted age groups is as expected, based on sampling strategy. If an important difference arises, check for errors (e.g. the incorrect school grades were included in the study, systematic errors in recording children date of births or date of measurement occurred, etc.). If possible, fix/correct potential errors. In any case, report the problem to WHO Regional Office (use the "Notes on COSI 2021-23 Datasets" template).

C. (M2) Child's age in month

Check the plausibility of values by taking into consideration the expected age of children and the following ranges of values for COSI targeted age groups:

- ✓ 6.0 – 6.9: 72.0 – 83.9 months
- ✓ 7.0 – 7.9: 84.0 – 95.9 months
- ✓ 8.0 – 8.9: 96.0 – 107.9 months
- ✓ 9.0 – 9.9: 108.0 – 119.9 months

Values out of these ranges can be either an error (data entry or recording error) or a younger or older child. In the latter case, the child will be excluded from the analysis (but he/she should be included in the dataset). Remember that an error in child's age will result in a wrong assessment of child's nutritional status, thus particular attention should be paid to this variable.

Based on children age in months, calculate age groups (6.0-6.9, 7.0-7.9, 8.0-8.9, 9.0-9.9 years old). Check that the number of children belonging to targeted age groups is as expected, based on sampling strategy. If an important difference arises, check for errors (e.g. the incorrect school grades were included in the study, systematic errors in recording children date of births or date of measurement occurred, etc.). If possible, fix/correct potential errors. In any case, report the problem to WHO Regional Office (use the "Notes on COSI 2021-23 Datasets" template).

D. (M3) Categorize the child's place of residence according to the country's urbanization grade

Provide the definitions/criteria that were applied for the three options "urban, semi-urban and rural" to WHO Regional Office (use the "Notes on COSI 2021-23 Datasets" template).

E. (O2) Child's place of residence, (O3) Child's postal code, (O4) Population size, (O5) Child region/municipality

Standardize each variable in order to have the same value for children with the same place of resident or postal code or region/municipality.

F. (M5) Date of measurement

- ✓ Check that values are consistent with the period during which the data collection was carried out.

- ✓ Since measurements are usually taken in a school within the same day or within few close days, check if there are children belonging to the same school who were measured within days of each other.

Remember that the date of measurement will be used to calculate children' age which is one of the parameters used to assess children' nutritional status. Thus, the date of measurement is a

very sensitive information that directly affects the quality of children' nutritional status assessment.

G. (O6) Hour/Minute of measurement

Check that the values are consistent with the opening hours of a primary school in your country.

H. Child's anthropometric measurements – (M8) Body weight, (M9) Body height, (O8) Second Height, (O9) Waist circumference and (O10) Hip circumference:

✓ Check that values are reported as follows:

- **Bodyweight expressed** in kilograms and recorded to the nearest 100-gram (0.1-kg) unit;

- **Body height expressed** in centimetres and recorded to the last completed 1 millimetre (mm) (0.1 cm).

- **Waist circumference** expressed in centimetres and recorded to the last completed 1 millimetre (mm) (0.1 cm)

- **Hip circumference** expressed in centimetres and recorded to the last completed 1 millimetre (mm) (0.1 cm).

If not, go for a more in-depth analysis in order to understand what happened and make an explanatory note to be sent to WHO Regional Office along with the dataset.

✓ Check if the dataset contains an excessive number of reported anthropometric measurements ending in particular digits, typically 0 or 5. If this is the case, reasons can be that the examiners rounded the measurements, or the instruments didn't work properly or didn't fit the requested technical features. Analyse the problem more in depth in order to understand what happened and make an explanatory note to be sent to WHO Regional Office along with the dataset.

✓ Check for extreme/implausible values analysing data by age groups.

✓ Calculate child's body mass index (BMI) and check for extreme/implausible values. BMI is calculated using the formula: weight (kg) divided by the square of the height (m^2).

I. (M10) Describe the clothes the child is wearing when measured

Check that this information is available for all children that have been measured. If the variable is blank, try to retrieve the information by checking the paper form and/or consulting with the examiner who took the measurements.

J. (M10) Describe the clothes the child is wearing when measured – [M10 – Other, specify]

Standardize values and translate them in English. If possible, provide the weight of the clothes reported in [M10 – Other, specify].

K. Observations by examiner

Read them carefully and assess if they somehow affect data quality or reliability. If yes, act accordingly: change variables values if applicable or report the inconsistency on the "Notes" variable. All observations that may be relevant at European level have to be translated in English.

Family's record form

A. Family's identification code (family's ID) – only if child's data and family's data are computerized separately (i.e. registered in two different datasets)

✓ Check that the family's ID was defined according to WHO 2021-23 data codebook - 13 digits as follows: country code (3 digits), the year of the data collection (2 digits), school identification code (4 digits), class grade (1 digit), class identification code (1 digit) and child's code (2 digits).

If a different definition was used, make a note to WHO Regional Office

B. (M1) What is your relationship to the child – "Other (please specify) I am _____"
Standardize values and translate them in English.

C. (O2) What is your child's date of birth?

✓ Check that the year of birth is compatible with the expected age of children. A common error that may occur is to register the date when the paper form was filled instead of child's date of birth.

✓ Check if the dataset contains an excessive number of children born on the same day or on the same month. If it is the case, make a more in-depth analysis in order to find a reasonable explanation (i.e. understanding if a data entry or recording error occurred).

D. (O4) What did your child weigh when he/she was born?
Check for values out of plausible range considering the information on full/preterm birth if available.

E. (M2) Was you child ever breastfed? – "Yes for _ _ months"
Check that the number of months is above 0. Check for values out of plausible range (range: 1-36, but it can vary from country to country according to local circumstances).

F. (O6) Was you child ever exclusively breastfed? – "Yes for _ _ months"

Check that the number of months is above 0. Check for values out of plausible range (range: 1-12, but it can vary from country to country according to local circumstances).

G. **(O10) if you chose the "Walking" or the "Cycling, skating or non-motorized scooter" answers above, how long does the journey usually take?**

Check that the number of minutes is above 0. Check for values out of plausible range (range: 1-60, but it can vary from country to country according to local circumstances).

H. **(M4) Over a typical week (including weekends), how much time on average per week does your child spend practising sports/exercise/dance on a sports club/health club/fitness centre/dance academy (e.g. football, track and field, hockey, swimming, tennis, basketball, gymnastics, ballet, fitness activities, dance classes, etc.)? - Hours**

Check for values out of plausible range (range: 0-15, but it can vary from country to country according to local circumstances).

I. **(M4) Over a typical week (including weekends), how much time on average per week does your child spend practising sports/exercise/dance on a sports club/health club/fitness centre/dance academy (e.g. football, track and field, hockey, swimming, tennis, basketball, gymnastics, ballet, fitness activities, dance classes, etc.)? - Minutes**

Check that the number of minutes is between 0 and 60.

J. **(M5) When does your child usually go to bed on school days?**

Check for values out of plausible range (range: 19:00-23:00, but it can vary from country to country according to local circumstances).

K. **(M6) When does your child usually wake up on school days?**

Check for values out of plausible range (range: 05:00-08:00, but it can vary from country to country according to local circumstances).

L. **(M7) Outside school hours, how much time on average per day does your child spend on playing actively at a moderate-vigorous intensity (e.g. running, jumping, playing non-supervised sports/dance, or physically active games)? – Hours**

Check for values out of plausible range (range: 0-6 for weekday and 0-8 for weekend day, but it can vary from country to country according to local circumstances).

M. **(M7) Outside school hours, how much time on average per day does your child spend on playing actively at a moderate-vigorous intensity (e.g. running, jumping, playing non-supervised sports/dance, or physically active games)? – Minutes**

Check that the number of minutes is between 0 and 60.

N. (O17) Outside school hours, how much time on average per day does your child do homework or read a book, either at home or somewhere else? – Hours

Check for values out of plausible range (range: 0-6 for weekday and 0-8 for weekend day, but it can vary from country to country according to local circumstances).

O. (O17) Outside school hours, how much time on average per day does your child do homework or read a book, either at home or somewhere else? – Minutes

Check that the number of minutes is between 0 and 60.

P. (M8) Outside school hours, how much time does your child on average per day watch TV or play with electronic devices (e.g. computer, tablet, smartphone; not including moving or fitness games)? – Hours

Check for values out of plausible range (range: 0-6 for weekday and 0-8 for weekend day, but it can vary from country to country according to local circumstances).

Q. (M8) Outside school hours, how much time does your child on average per day watch TV or play with electronic devices (e.g. computer, tablet, smartphone; not including moving or fitness games)? –Minutes

Check that the number of minutes is between 0 and 60.

R. (O26) We would also like to ask about your or your spouse's/partner's weight, height and age

Calculate BMI and check for extreme/implausible values. BMI is calculated using the formula: weight (kg) divided by square of the height (m^2).

As for age, check for values out of plausible range considering the information on the relationship to the child if available.

S. (M12) For the home where your child lives all or most of the time (>50%) please indicate the number of people, in each box, who live there – Someone else (please specify)

Standardize values and translate them in English.

T. (M12) For the home where your child lives all or most of the time (>50%) please indicate the number of people, in each box, who live there – Siblings

Check for values out of plausible range (range: 0-5 considering both brothers and sisters).

U. (O27) Was your child born in <insert country>?; (O28) Was the child's mother born in <insert country>?; (O29) Was the child's father born in <insert country>? – "No he/she was born in: _____"

Standardize values and translate them in English.

V. (O27a) If your child wasn´t born in , please indicate since when your child has been living here; (O28a) If your child´s mother wasn´t born in , please indicate since when she has been living here; (O29a) If your child´s father wasn´t born in , please indicate since when he has been living here – Month

Check that the value is between 1 and 12.

W. (O27a) If your child wasn´t born in , please indicate since when your child has been living here; (O28a) If your child´s mother wasn´t born in , please indicate since when she has been living here; (O29a) If your child´s father wasn´t born in , please indicate since when he has been living here – Year

Check for values out of plausible range taking into consideration the child's date of birth and/or mother/father's age if available.

X. (O30) In what language(s) do you usually/mainly speak with your child at home? – "Other language, please specify: _____"

Standardize values and translate them in English.

Y. Date of completion of this form

Check that values are consistent with the period during which the data collection was carried out.

Z. Remarks

Read them carefully and assess if they somehow affect data quality or reliability. If yes, act accordingly: change variables values if applicable or report the inconsistency on the "Notes" variable. All observations that may be relevant at the European level have to be translated in English.

School record form

A. School identification code (School ID)

Check that the school ID was defined according to WHO 2021-23 data codebook - 9 digits as follows: country code (3 digits), the year of the data collection (2 digits), school identification code (4 digits).

If a different definition was used, make a note to WHO Regional Office

B. (M1) What is your function at the school? – "Other (*please specify*)"

Standardize values and translate them in English.

C. (O4) What is the number of classes per grade selected in your school to participate?:

✓ For grades that were not included in the study: check that the variable is blank or equal to zero.

✓ For grades that were included in the study: check that values are consistent with the sample design adopted in the country. If the country followed the COSI Protocol, these variables should be equal to 1 – one class included per each grade; otherwise, values can be higher than 1 depending on country's choice.

D. (M2) For each participating class, please complete the columns below – "Class code" and "Grade/level of class"

✓ For codes: check that values are consistent with those used in child's identification codes (i.e. class identification code).

✓ For grades: check that values are consistent with grades included in the study.

E. (M6) In this current school year, for how much time each week does your school provide physical education lessons to the pupils of each class participating in this project?

Check for values out of plausible range (range: 0-180 minutes, but it can vary from country to country according to local circumstances).

F. Remarks

Read them carefully and assess if they somehow affect data quality or reliability. If yes, act accordingly: change variables values if applicable or report the inconsistency in the "Notes" variable. All observations that can be relevant at the European level have to be translated in English.

1.3 Additional information on procedures to check consistency between variables

Child's record form

A. (M3) Urbanization grade of the child's place of residence *versus* (O2) Child's place of residence, (O3) Child's postal code, (O5) Child region/municipality:

✓ Check that children with the same place of residence or postal code have the same value for the urbanization grade variable.

✓ Based on (O2) Child's place of residence or (O3) Child's postal code, check that the value of (M3) urbanization grade is correct.

B. (M6) Time of measurement *versus* (O6) Hour/Minute:

Check that (O6) Hour/minute value is coherent with the value of (M6) Time of measurement.

C. (M7) Child's Consent (Now I would like to weight you, measure your height and take your waist and hip circumferences. I will explain to you how I am going to do this. May I take these measurements?) *versus* **(O7) Refusal Reason (Can you tell me why you don't want to be measured?)**

Check that the following constraints are in place:
- ✓ If Child's consent variable is equal to "Yes", then Refusal Reason is blank
- ✓ If Refusal Reason variable is not blank, then Child's consent variable is equal to "No"

D. (M7) Child's Consent *versus* **Child's anthropometric measurements – (M8) Body weight, (M9) Body height, (O8) Second height measure, (O9) Waist circumference and (O19) Hip circumference**

Check that the following constraints are in place:
- ✓ If Child's consent variable is equal to "No" or it is blank, then all child's anthropometric measurements are blank.
- ✓ If at least one of the child's anthropometric measurements is not missed, then Child's consent variable is equal to "Yes".

Family's record form

A. (M2) Was your child ever breastfed? *versus* **Duration of breastfeeding in months [M2-Months]**

Check that the following constraints are in place:
- ✓ If (M2) is different from "Yes, for _ _ months", then [M2-Months] is blank
- ✓ If [M2-Months] is not blank, then (M2) is equal to "Yes, for _ _ months"

B. (O6) Was your child ever exclusively breastfed? *versus* **Duration of exclusive breastfeeding in months [O6-Months]**

Check that the following constraints are in place:
- ✓ If (O6) is different from "Yes, for _ _ months", then [O6-Months] is blank
- ✓ If [O6-Months] is not blank, then (O6) is equal to "Yes, for _ _ months"

C. (M2) Was your child ever breastfed? *versus* **(O6) Was your child ever exclusively breastfed?**

Check that the following constraints are in place:
- ✓ If (M2) is equal to "No", then (O6) is blank
- ✓ If (O6) is equal to "Yes for less than 1 month" or "Yes for _ _ months", then (M2) is different from "No"

- ✓ If (M2) is equal to "Yes, but for less than 1 month", than (O6) is different from "Yes, for _ _ months"
- ✓ If not blank, [M2-Months] is equal or greater than [O6-Months]

D. (O7) Was your child ever introduced to formula milk/infant formula? *versus* **((O7a) If yes, from which month was your child introduced to formula milk/infant formula?**

Check that the following constraints are in place:
- ✓ If (O7) is equal to "No", then (O7a) is blank
- ✓ If (O7a) is not blank, then (O7) is equal to "Yes"

E. (M3) How does your child usually get to and from school? *versus* **(O10) If you chose the "Walking" or the "Cycling, skating or non-motorized scooter" answers above, how long does the journey usually take?**

Check that the following constraints are in place:
- ✓ If [M3-To school] is not equal to "Walking" or "Cycling, skating or non-motorized scooter", then [O10-To school] is blank
- ✓ If [O10-To school] is not blank, than [M3-To school] is not equal to "School bus or public transport" or "Private motorized vehicles"
- ✓ If [M3-From school] is not equal to "Walking" or "Cycling, skating or non-motorized scooter", then [O10- From school] is blank
- ✓ If [O10- From school] is not blank, than [M3- From school] is not equal to "School bus or public transport" or "Private motorized vehicles"

F. (M3) How does your child usually get to and from school? *versus* **(O11) If your child doesn't walk or ride a bicycle, skateboard or non-motorized scooter from home to school or back home, please indicate the reason(s)**

Check that the following constraints are in place:
- ✓ If (M3 – both "To school" and "From schools") are equal to "Walking" or "Cycling, skating or non-motorized scooter", then (O11) is blank
- ✓ If (O11) is not blank, than (M3) is equal to "School bus or public transport" or "Private motorized vehicles" in at least one route

G. (M5) When does your child usually go to bed on school days? *versus* **(M6) When does your child usually wake up on school days?**

Calculate how many hours each child usually sleeps and check the plausibility of the values (range: 6 - 14 hours, but it can vary from country to country according to local circumstances).

H. (M7) Outside school hours, how much time on average per day does your child spend on playing actively at a moderate-vigorous intensity (e.g. running, jumping, playing

non-supervised sports/dance, or physically active games)?; (O17) Outside school hours, how much time on average per day does your child do homework or read a book, either at home or somewhere else?; (M8) Outside school hours, how much time does your child on average per day watch TV or play with electronic devices (e.g. computer, tablet, smartphone; not including moving or fitness games)?

Calculate the total amount of hours spent by each child usually playing actively at a moderate-vigorous intensity; doing homework or reading a book and watching TV or playing with electronic devices and check the plausibility of the values (range: 1 - 7 hours on a weekday and 1-12 hours on a weekend day, but it can vary from country to country according to local circumstances, e.g. the number of hours per day that the child stays at school).

I. **(O20) Does your child help to prepare family meals at home?** *versus* **(O21) If yes, please tell us about the food preparation activites that your child helps with at home**

Check that the following constraints are in place:
- ✓ If (O20) is equal to "No", then (O21) is blank
- ✓ If (O21) is not blank, then (O20) is equal to "Yes"

J. **(O27) Was your child born in <insert country>?; (O28) Was the child's mother born in <insert country>?; (O29) Was the child's father born in <insert country>?** *versus* **the country where the child/child's mother/child's father was born [O27/O28/O30 – Country]**

Check that the following constraints are in place:
- ✓ If (O27) is equal to "Yes", then [O27-Country] is blank. The same for (O28) and (O29).
- ✓ If [O27-Country] is not blank, then (O27) is equal to "No he/she was born in:_____". The same for (O28) and (O29).

K. **(O27) Was your child born in <insert country>?; (O28) Was the child's mother born in <insert country>?; (O29) Was the child's father born in <insert country>?** *versus* **If your child wasn´t born in <insert country>, please indicate since when your child/child's mother/child's father has been living here (O27a/O28a/O29a)**

Check that the following constraints are in place:
- ✓ If (O27) is equal to "Yes", then (O27a) is blank. The same for (O28 – O28a) and (O29 – O29a).
- ✓ If (O27a) is not blank, then (O27) is equal to "No he/she was born in:_____". The same for (O28a – O28) and (O29a – O29).

School record form

A. (M1) What is your function at school? *versus* Specification of other function [M1-Other function]

Check that the following constraints are in place:

- ✓ If (M1) is equal to "Headmaster/Headmistress/Principal" or "Teacher", then [4-Other function] is blank
- ✓ If [M1-Other function] is not blank, then (M1) is equal to "Other (please specify)".

B. (O4) What is the number of classes per grade selected in your school to participate? *versus* (M2) For each participating class, please complete the columns below

Check that the number of classes selected to participate is equal or greater than the number of classes reported in (M2), in total and by grade.

C. (O4) What is the number of classes per grade selected in your school to participate? *versus* (M6) In this current school year, for how much time each week does your school provide physical education lessons to the pupils of each class participating in this project?

Check that the number of classes selected to participate is equal or greater than the number of classes reported in (M6), in total and by grade.

D. (O4) What is the number of classes per grade selected in your school to participate? *versus* (M8) In this current school year, have any initiatives/projects been organized (or will be) in each class participating to promote a healthy lifestyle (e.g. to promote physical activity and/or healthy eating)?

Check that the number of classes selected to participate is equal or greater than the number of classes reported in (M8), in total and by grade.

E. (M2) For each participating class, please complete the columns below *versus* (M6) In this current school year, for how much time each week does your school provide physical education lessons to the pupils of each class participating in this project? *versus* (M8) In this current school year, have any initiatives/projects been organized (or will be) in each class participating to promote a healthy lifestyle (e.g. to promote physical activity and/or healthy eating)?

Check consistency between class identification code reported in (M2), (M6) and (M8).

F. (M2) For each participating class, please complete the columns below

For each gender, check that No. of pupils registered = No. of pupils examined + No. of pupils absent + No. of pupils who themselves declined to be examined + No. of pupils whose parents did not give the consent.

Please pay particular attention to the consistency of these values because they are used for the computation of children' participation rate which is an important indicator of data quality.

G. (M5) Does your school curriculum include physical education lessons? *versus* **Specify the grade or classes for which the school curriculum includes physical education [M5-Grade]**

Check that the following constraints are in place:
- ✓ If (M5) is not equal to "Only for some grade levels", then [M5 - Grade] is blank.
- ✓ If [M5 - Grade] is not blank, then (M5) is equal to "Only for some grade levels".

H. (M5) Does your school curriculum include physical education lessons *versus* **(M6) In this current school year, for how much time each week does your school provide physical education lessons to the pupils of each class participating in this process?**

Check that the following constraints are in place:
- ✓ If (M5) is equal to "No", then (M6) is blank.
- ✓ If (M6) is not blank, then (M5) is not equal to "No"

I. (M3) Does your school have outdoor playground area(s)? *versus* **(O5) Are the children allowed to actively play in extreme weather conditions (rain, snow, windy, hot) in outdoor playing areas?**

Check that the following constraints are in place:
- ✓ If (M3) is equal to "No", then (O5) is blank.
- ✓ If (O5) is equal to "Yes", then (M3) is equal to "Yes".

J. (M3) Does your school have outdoor playground area(s)? *versus* **(O6) Are children allowed to use outdoor playground areas outside school hours?**

Check that the following constraints are in place:
- ✓ If (M3) is equal to "No", then (O6) is blank.
- ✓ If (O6) is equal to "Yes", then (M3) is equal to "Yes".

K. (M4) Does your school have an indoor gym? *versus* **(O7) Are children allowed to use the indoor gym outside school hours?**

Check that the following constraints are in place:
- ✓ If (M4) is equal to "No", then (O7) is blank.
- ✓ If (O7) is equal to "Yes", then (M4) is equal to "Yes".

L. (O8) Does your school organize any sport/physical activities at least once a week for primary school children outside the school hours? *versus* **Specify the grade for**

which the school organizes some sport/physical activities at least once a week for outside the school hours [O8-Grade]

Check that the following constraints are in place:

✓ If (O8) is not equal to "Only for some grade levels", then [O8-Grade] is blank.

✓ If [O8-Grade] is not blank, then (O8) is equal to "Only for some grade levels".

M. (O8) Does your school organize any sport/physical activities at least once a week for primary school children outside the school hours? *versus* (O9) Do children attend these organized sport/ physical activities

Check that the following constraints are in place:

✓ If (O8) is equal to "No", then (O9) is blank.

✓ If (O9) is not blank, then (O8) is not equal to "No".

N. (O10) Is the school bus transport available to or provided by your school? *Versus* Specify the grade for which the school bus transport is available or provided [O10-Grade]

Check that the following constraints are in place:

✓ If (O10) is not equal to "Only for some grade levels", then [O10-Grade] is blank.

✓ If [O10-Grade] is not blank, then (O10) is equal to "Only for some grade levels".

O. (O10) Is the school bus transport available to or provided by your school? *Versus* Specify the distance for pupils living far away [O10-Distance]

Check that the following constraints are in place:

✓ If (O10) is not equal to "Only to pupils living far away", then [O10-Distance] is blank.

✓ If [O10-Distance] is not blank, then (O10) is equal to "Only to pupils living far away".

P. (M7) Does your school curriculum include nutrition education, either given as a separate lesson or integrated into other lessons? *versus* (O12) If yes, tell us what type of nutrition education your school provides

Check that the following constraints are in place:

✓ If (M7) is not equal to "Yes", then (O12) is blank.

✓ If (O12) is not blank, then (M7) is equal to "Yes".

2. Match the school, child and family datasets

2.1 Main steps

1) In case child's data and family's data are entered separately (i.e. in two different datasets): match each record in the children dataset with the corresponding record in the family dataset through the Child's ID and assess the level of matching.

   Take specific note of those children and families that do not match in order to exclude the occurrence of errors (data entry errors or errors in recording the ID code on paper forms). If no errors occurred, there are two possible scenarios for missed matching: a) parents gave the consent to their son/daughter's participation in the study but they did not fill in the Family record form; b) parents did not give the consent to their son/daughter's participation in the study but they filled in the Family record form. In the first case, child's data is available, and family's data is not; in the second case, it is the opposite.

2) After matching child's data with family's data, check that values of child's date of birth and gender reported by the examiner and the family are consistent, if possible. If an inconsistency arises, proceed as described in paragraph 1.1 - point 7.

   Consider that the reason of such an inconsistency could either be an error in recording the information (date of birth/gender), on paper or electronic form, or a matching error, meaning that family's information was assigned to the wrong child. The latter should be avoided because it will affect all findings on associations between children and families characteristics.

3) Match each record in the children dataset to the corresponding record in the school dataset through the School ID and assess the level of matching.

   Take specific note of those children and schools that do not match in order to exclude the occurrence of (data entry errors or errors in recording the ID code on paper forms). If no errors occurred, there are two possible scenarios for missed: a) the school headmaster allowed taking measurements at school premises but did not want to fill in the School record form; b) the school headmaster filled in the School record form but, for some reasons, measurements didn't take place in her/his school. Both scenarios are very unlikely, thus they should be carefully checked, and an explanation should be provided to WHO Regional Office.

3. Assess the nutritional status of children using the international growth reference curves (WHO and IOTF)

3.1 Main steps

1) Calculate the age of children in years or month (if not done before) as follows:
   - ✓ Age in years = (date of measurement-date of birth)/365.25
   - ✓ Age in months = (date of measurement-date of birth)/ 30.4375

2) Calculate the adjusted body weight of children by subtracting the weight of the clothes worn during the measurement. Each country should define an average weight of each type of clothing (i.e. underwear only, gym clothes, light clothing, heavy clothing, others). These values will compose the list of weight units for clothes to be used for adjusting children' weight.

3) Calculate child's anthropometric indicators and assess child's nutritional status according to 2007 WHO Reference for school-age children and adolescents (see paragraph 3.2).

4) Assess the prevalence of child overweight, obesity and thinness based on International Obesity Task Force (IOTF) Body mass index cut-offs (see paragraph 3.2).

3.2 Use the 2007 WHO Reference for school-age children and adolescents

The WHO Reference 2007 is a reconstruction of the 1977 National Center for Health Statistics (NCHS)/WHO reference. It uses the original NCHS data set supplemented with data from the WHO child growth standards sample for under-fives. To develop this reference the same statistical methodology was used as in the construction of the WHO standards.

The 2007 WHO recommended cut-offs for school age children and adolescents are used to compute height-for-age (H/A), weight-for-age (W/A) and BMI for-age (BMI/A) Z-scores, and to interpret anthropometric indicators:

- ✓ **BMI for-age -** Overweight, obesity and severe obesity are defined as the proportion of children with a BMI/A value above +1 Z-score, above +2 Z-scores and above +3 Z-scores, respectively, relative to the 2007 WHO growth reference median. Thinness and severe thinness are defined as the proportion of children with a BMI/A value below -2 Z-scores and below -3 Z-scores, respectively.

- ✓ **Height-for-age -** Stunting and severe stunting are defined as the proportion of children with an H/A value below -2 Z-scores and below -3 Z-scores, respectively, relative to the 2007 WHO growth reference median.
- ✓ **Weight-for-age -** Underweight and severe underweight are defined as the proportion of children with a W/A value below -2 Z-scores and below -3 Z-scores, respectively, relative to the 2007 WHO growth reference median.

According to WHO definitions, the prevalence estimates for stunted children include those who are severely stunted, the prevalence estimates for underweight children include those who are severely underweight, the prevalence estimates for thin children include those who are severely thin, and the prevalence estimates for children with overweight include those who are with obesity.

Children with biologically extreme/implausible anthropometric indicators values are excluded from the analysis. Extreme or biologically implausible z-scores are defined as follows:
- ✓ W/A Z-score below -6 or above 5
- ✓ H/A Z-score below < -6 or above 6
- ✓ BMI/A Z-score below < -5 or above 5

Two WHO tools for applying 2007 WHO growth reference are available at https://www.who.int/toolkits/growth-reference-data-for-5to19-years/application-tools:
1. WHO AnthroPlus software (Nutritional Survey module)
2. Macros in R, SAS, S-Plus, SPSS and Stata

WHO Reference tables and charts are available at https://www.who.int/tools/growth-reference-data-for-5to19-years/indicators.

Both WHO AntrhoPlus software - Nutritional Survey module and Macros perform the following:
- ✓ Calculation of the Z-scores for the three anthropometric indicators (weight-for-age, height-for-age and BMI-for-age).
- ✓ Identification of extreme z-scores (i.e. biologically implausible) for each indicator.
- ✓ Production of sex- and age-specific estimates for the prevalence of under/over nutrition and summary statistics (mean and SD) of the z-scores for each indicator using all available (non-missing and non-flagged) z-score values.

Before downloading and running the Macro read carefully the instructions provided at https://www.who.int/toolkits/growth-reference-data-for-5to19-years/application-tools.

Before downloading and installing the WHO AntrhoPlus software read carefully the Manual provided at https://www.who.int/toolkits/growth-reference-data-for-5to19-years/application-tools.

Both WHO AntrhoPlus software and Macros allow considering (with some limitations) the sample nature of data when calculating prevalence and confidence intervals. Nevertheless, it is highly recommended checking that all aspects of the applied sampling strategy (sample weights, stratification and clustering) can be properly taken into consideration by the software/Macro. If not, it is better to produce the sex- and age-specific estimates for the prevalence of under/over nutrition and summary statistics without using these tools. The prevalence values can be estimated using a common statistical software (SPSS, Stata, SAS, R) after creating the following variables that are based on height-for-age (H/A), weight-for-age (W/A) and BMI for-age (BMI/A) Z-scores:

- ✓ Thinness: 1 if BMI/A Z-scores < -2; 0 if not
- ✓ Overweight: 1 if BMI/A Z-scores > +1; 0 if not
- ✓ Obesity: 1 if BMI/A Z-scores > +2; 0 if not
- ✓ Stunted: 1 if H/A Z-scores < -2; 0 if not
- ✓ Underweight: 1 if W/A Z-scores < -2; 0 if not.

3.3 Use the International Obesity Task Force (IOTF) Body mass index cut-offs

Children aged 2 to 18 are classified as thin, normal weight, with overweight or with obesity based on age-specific (in months) and gender-specific BMI cut-offs. These cut-offs are based on and linked to the corresponding adult BMI cut-offs. More specifically, they are available corresponding to the following body mass index (BMI) cut-offs at 18 years:

- ✓ 16 kg/m^2 thinness grade 3
- ✓ 17 kg/m^2 thinness grade 1
- ✓ 18.5 kg/m^2 thinness grade 1
- ✓ 23 kg/m^2 overweight (unofficial Asian cut-off)
- ✓ 25 kg/m^2 overweight
- ✓ 27 kg/m^2 obesity (unofficial Asian cut-off)
- ✓ 30 kg/m^2 obesity
- ✓ 35 kg/m^2 morbid obesity:

Observed BMI is calculated using the formula: adjusted body weight (kg) divided by height squared (m^2) and it is compared to the BMI cut-offs for children of the same gender and age in months.

The prevalence values of overweight and obesity can be estimated using a common statistical software (SPSS, Stata, SAS, R) after creating the following variables:

- ✓ Overweight: 1 if observed BMI >= corresponding BMI cut-offs for overweight; 0 if not
- ✓ Obesity: 1 if observed BMI >= corresponding BMI cut-offs for obesity; 0 if not.

According to these definitions, the prevalence estimates for children with overweight include those who have obesity.

4. Calculate indicators to assess data quality

The following indicators should be calculated:

1) <u>Participation in the study</u>, measured as follows:

1. No. of Sampling Units that participated in the study / No. of Sampling Units that were invited to participate * 100 (Sampling Units Participation Rate).

   The rate is calculated for each stage included in the sample design. Reasons for refusal should be reported if available.

2. No. of children who were examined / No. of children who were invited to participate * 100 (Children Participation Rate).

   The rate is calculated by gender and for all children.

3. No. of children who refused themselves to be measured / No. of children who were invited to participate * 100 (Children Refusal Rate).

   The rate is calculated by gender and for all children. Reasons for refusal should be reported.

4. No. of children whose parents didn't give the consent to measurement/ No. of children who were invited to participate * 100 (Parents Refusal Rate).

   The rate is calculated by gender and for all children.

5. No. of children absent the day of the measurement / No. of children who were invited to participate * 100 (Children Absence Rate).

   The rate is calculated by gender and for all children.

2) <u>Completeness of information needed to calculate children's anthropometric indicators</u>, measured as follows: No. of children with complete information on age, gender, weight, height, clothes worn during the measurement / No. of children with consent to be measured * 100.

3) <u>Matching between children, family and school datasets</u>, measured as follows:

1. No. of children who were examined and with data collected through the Family's record form / No. of children who were examined * 100.

2. No. of children who were examined and with data collected through the School record form / No. if children who were examined * 100.

5. Create the datasets as requested by WHO Regional Office

Countries that did not use LibreClinica to enter and store the data collected through COSI paper forms, are kindly request to send to WHO Regional Office the following two datasets:

1) Children dataset, which contains all data collected through the Child's record form and the Family's record form

2) School dataset, which contains all data collected through the School record form.

Individual sampling weights to adjust for the sampling design and sampling variables (e.g. sampling units ID, stratification variables, etc.) have to be added to children and school datasets, if available at country level.

Format, values and labels of all variables in both datasets should be consistent with the COSI 2021-23 Data Codebook to the greater extent.

All deviations from the Codebook and from COSI Record forms have to be reported to WHO Regional Office using "Notes on COSI 2021-23 Datasets" template.

Countries are also kindly requested to provide WHO Regional Office with the following information, if relevant:

- ✓ A brief description of how the levels of urbanization grade used in COSI 2021-23 child's form were defined - including if they have been defined based on the school location or on the child's residence place.
- ✓ For each level of education used in COSI 2021-23 family form, the number of school years needed to complete the level according to the national education system.
- ✓ The list of weight units for the given options of the clothes worn by the children when measured (underwear only, gym clothes, light clothing, heavy clothing).
- ✓ A brief description of sampling weights calculation if relevant.

The above-mentioned information should be reported to WHO Regional Office using "Notes on COSI 2021-23 Datasets" template. If the list of weight units cannot be provided, both the non-adjusted and adjusted body weight should be reported in children dataset.