

SOP 0645. Checks to be carried out for Datasets before producing tables

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>

1. What

This SOP describes the procedure for the final checking for consistency and completeness of any dietary assessment projects coded and analysed using DINO.

2. When

To be carried out on completion of diet coding at the end of the study, enabling production of the dietary datasets

3. Who

To be used by anyone who is involved in using DINO for diet coding and data analysis.

4. Why

Dietary data checks are required prior to dataset production to;

- Reconcile any discrepancies in the data.
- Ensure there is no missing data.
- Check the feasibility of the data and portion sizes.

5. Procedure

There are a number of data checks to be carried out before producing dietary datasets. This document will explain each data check required and how to perform the check, together with instructions on making amendments if necessary. Where possible, it is advisable to begin these checks well in advance of when the final dataset is due. The checks can then be run again at a later date to confirm there are no outstanding discrepancies. All checks must be completed before dietary datasets are produced.

The checks are split into two phases (phase 1 must be completed prior to phase 2):

Note: Before starting you should check that all food records have been coded and related food queries have been edited.

Phase 1

- 6.1 Unknown food or recipe codes
- 6.2 Number of days of recording and correct subject numbers
- 6.3 Missing 'Where/With whom' data

Phase 2

- 6.4 Feasibility checks for all reported foods (portion sizes)
- 6.5 Extreme intake checks for all reported nutrients

Phase 1

6.1 Unknown food or recipe codes

Unknown food codes are often used as temporary codes while queries are raised and resolved. This check ensures that all unknown food codes have been deleted or replaced with known food codes.

SOP 0645. Checks to be carried out for Datasets before producing tables

1. From DINO's main switchboard, click '**Data Analysis**', then '**Subjects eating particular foods**'. Select the appropriate study from the drop down menu. Go to '**Foods**', then either:
 - click 'Select from list.' Highlight food code '99999 Unknown/ Recipe Food', then click 'Run Analysis'.
 - click 'enter a single food code', enter food code 99999, then click 'Run Analysis'

If there are no respondents on the list, there are no unknown foods in the dataset

2. Subject ID's shown on the list are those with unknown foods still coded. Along with subject ID's, the list also shows the food record dates, and grams and portions of unknown food. Go through the list and for each respondent using the '**Review coding**' form in '**Dietary Coding**' and the food record if necessary to amend these entries. Some entries will require deletion and others will need updating with the appropriate food code.
3. Once entries have been amended for each respondent, repeat step 1 to ensure all unknown foods have been dealt with.

6.2 Number of food record days and correct subject numbers

This check is to ensure that the number of days coded matches the number of days that were completed in the food record.

1. From DINO's main switchboard, click '**Data Analysis**', then '**Subject Diary records**'. Select the appropriate study from the drop down menu and click Run Analysis. The output can be produced in 'tabulation' or in 'excel'.
2. The output will show all the subject IDs applicable to that study and a column called No. of days span, which is the number of different days the coding spans across. By eye, scroll down this list to check that the days coded matches the days as expected.
3. Any discrepancies could be due to a number of reasons; an error in the entry of dates in the coding form, or a food record day that has not been entered due to insufficient information.
4. Go through the list, noting the reasons for these discrepancies and making any necessary corrections to the entries in DINO, via the review coding form.
5. This record can also be used to check that the number of Subject ID's coded matches the number of food records expected to be coded. If you do this via 'tabulation' you can click on the little box on the scroll bar which will tell you the number of records coded. It may be beneficial to run the analysis in Excel so that you can apply a count, instead of manually counting up the ID's.

6.3 Missing where/with whom data

This check is to ensure that the 'where?/with whom?/TV on?/At table?' questions are complete for each food/drink item coded.

Note: *These questions are not be required for all studies.*

SOP 0645. Checks to be carried out for Datasets before producing tables

1. From DINO's main switchboard, click '**Data Analysis**', then '**analyse-with final data export**'. Create a blank record, name the analysis and enter your initials before selecting the study. Do not select any nutrients, simply check the '**where etc.**' box at the bottom of the form. Run the output to excel and save the document. A pop up will inform you when the export has been done. Sort the data so that you can identify any missing questions.

The list shows all food/drink items for which there is missing 'where?/with whom?/TV on?/At table?' data, along with the date and time of consumption and the subject ID. Use the '**Review Coding**' form in **DINO** to find each food item, and then click 'Amend Record'. Go to the 'Questions' tab and using the respondents food record, fill in any missing information.

When the phase 1 checks have been completed, progress to phase 2. If multiple amendments have been made following phase 1 it may be necessary to run some outputs again to ensure other errors were not previously masked.

Note: There are further checks that can be carried out. More information can be found in DINO. In Dino click on Main menu, Dietary coding, coding reports, Phase 1 checks.

Phase 2

6.4 Feasibility checks for all reported foods

This check is to review the total amount of foods consumed by each respondent, by daily intake, allowing you to see extreme portion sizes (very large or very small) by eating occasion. Thereby ensuring data entry errors in portion sizes are checked and amended where necessary.

1. From DINO's main switchboard, click '**Data Analysis**', then '**analyse-with final data export**' and create a new record. Label this record **Phase2_feasibilitychecksfoods** (date) and place initials in the author box. Select the study required.
2. Do **NOT** select any nutrients or food groups, this will ensure all foods and food groups are produced in the output, together with a gram value. Select 'tabulation' to check the output will be as desired then select Excel to export the data into a spread sheet in the chosen destination. A pop-up will inform you when the export has been done. In large studies this output may take 2-3 minutes to produce. Ensure the option of Excel 2007-10 is selected (xlsx) to give unrestricted rows of data.
3. The export/output should have the following headers;
Study title, Subject ID, Gender, Date of Birth, Diary date, Day of week, Meal time description, Meal time, Food name, Main Food Group code, Main Food Group description, Sub Food Group code, Sub Food Group description, Recipe name, Recipe main food group code, Recipe Main Food group description, Recipe sub Food Group, Total grams (as shown below).

SOP 0645. Checks to be carried out for Datasets before producing tables

Study Title	Subject ID	Subject Ref	Gender	DateOfBirth	DiaryDate	Day of Week	DayNo	MealTimeD	MealTime	FoodName	MainFoodGroupCode	MainFoodGroupDescription	SubFoodGroupCode	SubFoodGroupDescription	RecipeName	RecipeMainFoodGroupCode	RecipeMainFoodGroupDescription	RecipeSubFoodGroupCode	RecipeSubFoodGroupDescription	Total Grams
1																				

- In excel, highlight all the cells and sort the data.
Sort the spread sheet by Main Food Group description (A – Z), then by Food Name (A-Z), then by Total Grams descending, smallest to largest. See below.

The image shows the 'Sort' dialog box in Excel. At the top, there are buttons for 'Add Level', 'Delete Level', and 'Copy Level', along with an 'Options...' button and a checkbox for 'My data has headers' which is checked. The main area has three rows for sorting criteria:

Column	Sort On	Order
Sort by	MainFoodGroupDes	Values (A to Z)
Then by	FoodName	Values (A to Z)
Then by	Total Grams	Values (Smallest to Largest)

At the bottom are 'OK' and 'Cancel' buttons.

If there are multiple age groups in the study you may want to sort the spreadsheet by age, by cutting and pasting the entries into separate tabs for each age group.

- In excel, highlight all the cells, under the data tab select the subtotal function and complete the pop up box as shown here, selecting the Min function.

The image shows the 'Subtotal' dialog box in Excel. It has the following settings:

- At each change in:** Recipe Main Food Group
- Use function:** Min
- Add subtotal to:**
 - ☐ FoodName
 - ☐ Main Food Group
 - ☐ Sub Food Group
 - ☐ Recipe Main Food Group
 - ☐ Recipe Sub Food Group
 - ☒ Total Grams
- ☒ Replace current subtotals
- ☐ Page break between groups
- ☒ Summary below data

At the bottom are 'Remove All', 'OK', and 'Cancel' buttons.

Repeat exactly the same a second time **but** select the Max function and uncheck the replace current subtotals box;

SOP 0645. Checks to be carried out for Datasets before producing tables

6. Is it recommended that this check is carried out at the food group level, but this can be unmanageable in large datasets. If checking by food repeat as above under 'at each change in' choose 'food name'.
7. Excel will display the minimum and maximum portion size consumed within the food group (or food name). Clicking on the 1 in the top left corner will compress all levels of the data, the 4 will expand the data by food group allowing you to see all entries.
8. Look for any obvious anomalies in the minimum and maximum portion sizes e.g. 900g maximum of biscuits or 0.1g of porridge. When you notice something unusual expand the food group and identify the food in the food name column and subject ID. Check review coding in DINO for clues to the unusual portion and use the food record to correct if appropriate.
9. Where errors in data entry are noticed you must write this in the spread sheet, alongside the original entry, stating the correction made, date and initials. If the portion size is truly high or low then no corrections should be made, but the entry should still be marked in the spread sheet as checked. Once the corrections are made the next check (6.5) can be carried out.

6.5 Extreme intake checks for all reported nutrients

This check should be completed after the feasibility checks. This check reviews the daily nutrient intakes for each respondent allowing you to see extreme cases of nutrient intakes. This may flag errors in data entry or errors in the food composition data.

1. From DINO's main switchboard, click '**Data Analysis**', then '**analyse-with final data export**' and create a new record. Select the study, give a record name and initial as previous.
2. Within the nutrient codes tab select all nutrients which are required in the final dataset. To do this click and drag to create a list or hold Ctrl while clicking on each nutrient.
3. Select daily totals, but do not select any food groups, this will provide an output of data across all foods. Check the excel box and press test to export to excel. A pop-up will inform you when the export has been done.
4. Repeat stage 2 and 3 for each individual nutrient that you are checking.
5. Within the daily totals output spreadsheet, highlight all cells and sort by chosen nutrient (largest to smallest). Identify and highlight which values need checking and make a note of the subject ID and diary date/day number. Go to the relevant individual nutrient spreadsheet and search by subject ID and diary date, to see which foods are contributing to the high values.
6. Review all nutrients for very high and very low values. Cross-reference with the feasibility checks and the food record to ensure accurate data entry and amend using review coding if necessary. Where amendments are made these should be clearly described in the spread sheet for reference.
7. There are no set numbers/boundaries to use when checking nutrient intakes and it can be done by eye by those who have an awareness of nutrient data and recommended intakes.

SOP 0645. Checks to be carried out for Datasets before producing tables

8. When the data checks have been completed the dietary datasets can be produced. See SOP 6 – Dietary Dataset production using DINO.

6. Related Documentation

Doc Number	Title
SOP 6	Dietary Dataset production using DINO