

FAQ EPIC Nutritional Data

1. Introduction to the EPIC study

The EPIC study gathers data from several European cohorts.

Dietary assessments tools

Diet was assessed at study baseline using validated country/centre-specific methods, including dietary questionnaires (DQs) spanning the previous 12 months. In most centres, DQs were self-administered, except for Ragusa (Italy), Naples (Italy), and Spain, where face-to-face interviews were performed. Extensive quantitative DQs were used in northern Italy and Germany and were structured by meals in Spain, France, and Ragusa (Italy). Semi-quantitative food-frequency questionnaires (FFQs) were used in Denmark, the Netherlands, Norway, Naples (Italy), Umeå (Sweden), and the United Kingdom, and a FFQ was combined with a 7-day record on hot meals in Malmö (Sweden). Please refer to the document [List of available dietary variables and EPIC questionnaires](#) in the sheet **Baseline diet assessment**.

Follow-up DQs have been sent in some centres. Please refer to the document [Data collection dates – Dietary by Centre](#). Centralization of this data at IARC has been planned but has not yet been effectively implemented.

A common standardized dietary assessment method, a single 24-hour dietary recall, has been applied to all EPIC centres on a subset of the cohort (about 37 000 subjects) to enable calibration of the data. Data were collected using a highly standardized tool named EPIC-Soft (later renamed GloboDiet) developed at IARC in collaboration with the EPIC centres between March 1995 and June 2000.

Both DQ and 24-hour dietary recall data were classified using the EPIC-Soft food classification, ECLASS (see [question 4](#)).

Dietary components

To study nutrient intakes, the EPIC Nutrient Database (ENDB) was compiled using the calibration data, involving the EPIC collaborators and local food composition table compilers. This harmonized database was linked to the DQ items at a later stage. A first release of nutrient data occurred in 2009 for DQs and 24-hour dietary recall data. Folate intakes were added shortly thereafter, and then methyl-group donors were added. To substantially increase the number of nutrients, EPIC data were matched to the United States Department of Agriculture (USDA) Nutrient Databases and nutrient intakes released in 2018.

In parallel, to study contaminants and other diet-related variables (e.g. greenhouse gas emissions, land use), the EPIC data were classified using the European Food Safety Authority (EFSA) FoodEx and FoodEx2 classification systems, and also by NOVA food processing groups.

Other components were also studied. For the full list, please refer to the document [Food components](#).

2. Which manuscripts are used as references for the baseline dietary assessment data?

Overall description of the baseline dietary questionnaires and their validation

Riboli et al. 2002 ([PMID: 12639222](#)) remains a suitable reference for the overall study design but is outdated concerning the description of dietary assessment methods. For descriptions and information on the validation of dietary tools in each country/centre, please refer to the document [List of available dietary variables and EPIC questionnaires](#) in the sheet **Baseline diet assessment**. The reference is **Huybrechts et al. 2022** ([PMID: 36590209](#)).

When using FoodEx- or FoodEx2-derived data (contaminants, biodiversity, SHARP):

Please refer to **Huybrechts et al. 2025** ([PMID: 40460259](#)).

When using NOVA classification/processed foods data:

Please refer to **Huybrechts et al. 2022** ([PMID: 36590209](#); Erratum in [PMID: 37260516](#)).

A supplement of the *International Journal of Epidemiology* was dedicated to the validation of EPIC dietary assessment methods in eight countries in 1997 – Umeå (Sweden), Norway, and Denmark are not included in this publication ([PMID: 9126528](#)). We therefore recommend using the updated versions ([PMID: 36590209](#)).

Overall description of the calibration data collection and validation of the method

If you studied 24-hour dietary recall data alone or if the calibration approach was used, then please refer to the collection of calibration data mentioning **Slimani et al. 2002** ([PMID: 12639223](#)). In 2009, volume 63, supplement 4 of the *European Journal of Clinical Nutrition* gathered several publications on nutrient intake and patterns of the calibration data.

Validation papers include [PMID: 19056549](#) (plasma fatty acids) and [PMID: 16160701](#) (plasma carotenoids).

3. Which manuscripts are used as references for sources of nutrients/dietary components, indices, scores, and processing codes?

The document [List of available dietary variables and EPIC questionnaires](#), in the sheet **List of variables**, lists the reference manuscripts.

4. The EPIC classification – How were the baseline FFQ recipes and dishes decomposed in EPIC?

As stated in the introduction, the different centres/cohorts that joined EPIC used different dietary assessment methods. A calibration method, namely a single 24-hour dietary recall collected by trained interviewers on a subsample of each centre/cohort, was set up using the EPIC-Soft application.

The EPIC classification is the one adopted for collecting the 24-hour dietary recalls with EPIC-Soft. The EPIC-Soft food classification was developed by IARC and the EPIC centres ([PMID: 12082516](#)). In the early 1990s, the EPIC team was largely inspired by ongoing work within the European Union (EU) COST Action 99/Eurofoods project, which was building the Eurocode 2 classification and the LINGUAL description systems. The Eurocode 2 Food Coding System was originally developed within the European FLAIR Eurofoods-Enfant Project “to serve as a standard instrument for nutritional surveys in Europe and to serve the need for food intake comparisons” ([PMID: 1486874](#)). Because the LINGUAL thesaurus is very detailed to describe food samples included in food composition tables (like the FoodEx2 system is now), only the concept of facets and descriptors was kept and simplified to the level of detail expected to be recalled by lay cohort subjects.

The EPIC-Soft classification system used for the 24-hour dietary recall data ([PMID: 12082516](#)) was applied to the dietary questionnaire data. This implied decomposing dishes, such as stews and sandwiches, like they were handled in EPIC-Soft. Each country provided standard recipes for their dishes.

The presentation [2022_03_18_Mother-Daughter_Matrix_use.pdf](#) explains how FFQ food items were handled. It can be obtained upon request to epic-diet@iarc.who.int.

5. In terms of classification, what level of detail can be used in the analyses?

Because of differences across dietary assessment tools, the level of details differs between countries. For certain foods or food components it may be appropriate to restrict the analyses to the countries that provide the required level of detail.

Please refer to the document [Food classifications](#). The sheet **README** provides an overview of the different worksheets, explaining the content and/or meaning of variables.

The sheet **Food Classification (ECLASS)** provides guidance on the classification level to be used for analysis. A common good practice would be:

- 1- to start with ECLASS subgroups because they are more harmonized across centres,
- 2- to choose NCLASS when more details are needed, and
- 3- to look at the *selected items*, as an alternative to more detailed data.

Please always check data availability per centre, for dietary questionnaires or 24-hour dietary recalls, provided in dedicated sheets.

When working on **diet index/score** composition, component availability should be checked for each centre and then also for the overall index/score, because availability in the considered centres may differ across the components, thus influencing the overall centres retained for the overall index calculation.

Be careful when mixing ECLASS and NCLASS data; the links between the classifications are provided in the sheet **Classification Linkage**.

Descriptive information on baseline consumption data is also provided as an aid, as mentioned in [question 13 “What are the intakes of food groups and subgroups and of nutrients in EPIC?”](#).

Please be aware that to ease intake handling in programs, missing quantities were replaced with zeroes. Availability of variables per centre is indicated in dedicated sheets in the document [Food classifications](#).

Examples of specific groupings

Based on the detailed classification, dietary variables can be built for specific analysis purposes, especially to build diet scores. A few examples are given here.

Wholegrains

In EPIC, we have distinguished between “white” (refined) and “non-white” (which includes mixtures of refined and non-refined ingredients, as well as semi-refined and wholegrain products). This is available for pasta, rice, breads, and crispbreads in the NCLASS system only, for some centres and sometimes based on assumptions about local habits (i.e. the level of detail was not asked in the questionnaire).

“Wholegrains” is a component of many scores built to assess the healthiness of diet. We advise considering some NCLASS groups and conversion factors as described in the sheet **Wholegrains** in the document [Food classifications](#).

High fat versus low fat dairy products

In EPIC, milks, milks for coffee, fermented milks, yogurts, and curds (but not cheese, desserts, and creams) are classified as skimmed, semi-skimmed, and whole (or non-specified) in the NCLASS system only. Note that we used the European standards to classify milks: “whole” contains 3% fat, and “semi-skimmed” contains about 1.5% fat.

The sheet **Dairy products** in the document [Food classifications](#) provides examples of groupings made to produce the variables such as those shown below, available for some centres:

GRA_LFATDAIRY	Low fat content dairy products
GRA_HFATDAIRY	High fat content dairy products
DAIRY_FAT_SSK	Group of semi-skimmed dairy products
DAIRY_FAT_SK	Group of skimmed dairy products
DAIRY_FAT_WF	Group of whole fat dairy products
DAIRY_FAT_NS	Group of dairy products without fat content information

6. Which nutrients and other compounds are available?

Please refer to the document [Food components](#).

Definitions of nutrients from the EPIC Nutrient Database (ENDB) and the USDA Database (USDA) are provided in the sheets **ENDB** and **USDA**, respectively.

The sheet **Guidance US** contains the guidelines to use the extended (US) EPIC nutrient database; it provides some guidance for the use of amino acids and fatty acids.

Depending on your analysis aims, intakes of compounds can be provided globally (i.e. for all food groups combined) or only for given food groups or subgroups (e.g. iron from red meats only).

NB: Do not mix ENDB and USDA nutrients; use one set or the other. As a rule of thumb, we recommend using the ENDB for nutrients available in both ENDB and USDA. However, if you need nutrients that are not available in the ENDB, then we recommend you use the USDA for all the

nutrients included in your analyses. ENDB and USDA nutrients should not be mixed because some of the nutrients may be related to one another and/or derived from each other (e.g. energy is the sum of energy derived from macronutrients).

For contaminants, the original data were provided by the European Food Safety Authority (EFSA). *Cadmium* intakes are available, but it should be acknowledged that diet is not the only source of cadmium, because tobacco smoking is also a prevalent source. To ease the work on persistent organic pollutants (POPs), guidelines are currently being prepared. We are waiting for a validation paper for mycotoxins before disseminating the corresponding data.

7. Plant versus animal origin

During the ENDB project, an attempt was made to classify the foods according to their source, plant versus animal: the variable ORIGIN was generated. This enables us to provide the **origins of the nutrients Fat and Protein**.

Details on how this task was done are given in the worksheet **ORIGIN** in the document [Food components](#).

Variable name	Variable label	Type
QE_ENERGY	DQ ENDB Energy (kcal)	Energy
QE_WATER	DQ ENDB Water (g)	Proximates
QE_PROT	DQ ENDB Total proteins (g)	Proximates
QE_PROTPL	DQ ENDB Total proteins, plant origin (g)	Proximates
QE_PROTAN	DQ ENDB Total proteins, animal origin (g)	Proximates
QE_PROTUK	DQ ENDB Total proteins, unknown origin (g)	Proximates
QE_FAT	DQ ENDB Total fats (g)	Proximates
QE_FATPL	DQ ENDB Total fats, plant origin (g)	Proximates
QE_FATAN	DQ ENDB Total fats, animal origin (g)	Proximates
QE_FATUK	DQ ENDB Total fats, unknown origin (g)	Proximates
QE_CHO	DQ ENDB Carbohydrates (g)	Proximates
QE_SUGAR	DQ ENDB Sugar (g)	Proximates
QE_STARCH	DQ ENDB Starch (g)	Proximates
QE_FIBT	DQ ENDB Total dietary fibre (g)	Proximates
QE_ALC	DQ ENDB Alcohol (g)	Proximates

8. Which indices, scores, and processing codes are available?

Please be aware that existing pre-calculated diet scores can be used as such for adjustments. If a score is a main research variable, it will have to be calculated based on the centres included in the project if cut-offs are population-based (e.g. use of medians or quintiles).

Several indices and scores have been prepared at IARC: see the documents [List of available dietary variables and EPIC questionnaires](#) and [Scores – Processed foods](#).

Some other indices or scores have also been calculated outside IARC, for which scripts/codes are stored at IARC. These can be shared with researchers who wish to adapt them to their data and/or research questions.

Please contact epic-diet@iarc.who.int if you wish **to access** the scripts or **to share** the scripts of a new indicator with the EPIC community.

At the stage of the Data Transfer/User Agreement, the user can request ECLASS, NCLASS components, nutrients, etc. as needed, as well as pre-calculated diet scores.

Mediterranean diet scores

Since the *Original Mediterranean diet score* developed by Trichopoulou et al. 1995, a number of variants of this score have been prepared, as explained in the sheet **MED scores** in the document [Scores – Processed foods](#).

Some scores have been calculated outside IARC, for which IARC may not have the scripts.

IARC disseminates primarily the recent MED scores published by Romaguera et al. 2010 (PMID: 20810975), namely the relative MED score and the modified relative MED score.

Plant-based diet index

The plant-based diet index has been adapted to EPIC data from the original publication by Satija et al. 2017 (PMID: 28728684) as described in Shah et al. 2025 (PMID: 40658296). The overall, healthful, and unhealthful plant-based diet indices have been calculated for baseline questionnaire data.

Classifications included in the EPIC plant-based index are provided in the sheet **Classification PDI** in the document [Scores – Processed foods](#).

NOVA processing codes

The sheet **NOVA codes** in the document [Scores – Processed foods](#) provides the subgroup codes used in EPIC, together with some examples and comments. Subgroups can be used in analyses to study a given dietary group within a NOVA group (e.g. dairy products in Nova 3 Processed foods).

The sheet **NOVA groups** shows in which NOVA group foods fall, organized by food group, for example foods classified in the food group Meat (see supplementary Table 1 in PMID: 39867840). This enables viewing the distribution of, for example, meat items into the different NOVA groups.

Please be aware that Water (Nova code 101) is not asked in all countries, so when studying Nova 1 in g/day, it would be a good idea to run sensitivity analyses for Nova 1 without the Water food group.

When using NOVA classification/processed foods data:

Please refer to Huybrechts et al. 2022 (PMID: 36590209; Erratum in PMID: 37260516).

Dietary inflammatory indices

There are two dietary inflammatory indices available in EPIC; both have been compared with inflammatory biomarkers in Lécuyer et al. 2023 (PMID: 37271707).

The **Dietary Inflammatory Index (DII)** (PMID: 23941862) is the original index developed by James R. Hébert and Nitin Shivappa. Three indices have been calculated: not adjusted for energy, adjusted for energy using the residual approach, and adjusted for energy using the density approach. Data were calculated once for all.

A modified version of this DII was developed by the team of Antonio Agudo (PMID: 29635497) within EPIC, named the **Inflammatory Score of the Diet (ISD)**. A version of the ISD without alcohol is available, as well as data by quartiles.

9. When do/can we use the calibration approach?

Guidelines are in preparation. It should be noted that the calibration method can only be applied when the dietary variables of interest are available for both the FFQ and 24-hour dietary recall data (e.g. there is no food contaminant data available for the 24-hour dietary recalls). In addition, the distribution of the variables of interest should be checked, because the calibration approach may not be applicable when distributions are too skewed (e.g. for food components and contaminants that have a large number of zero values/intakes).

10. Availability of dietary supplements

Calibration data: The intake of dietary supplements was asked in the EPIC-Soft 24-hour dietary recall interview. Please refer to **Skeie et al. 2009** ([PMID: 19888276](#)) to know what supplements were reported on the recalled day.

Baseline data: The intake of dietary supplements was not asked in all centres (asked in Denmark, Malmö (Sweden) menu book, Umeå (Sweden), the Netherlands, Norway, and the United Kingdom with details, in Germany with yes/no). When it exists, this information has not been centralized at IARC.

11. Exclusion criteria

Not all EPIC subjects are retained when running statistical analyses; this was discussed and agreed in statistical working group meetings in the early 2000s. Eventually a reference to this conclusion was included in **Ferrari et al. 2008** ([PMID: 18180242](#)), which comprehensively described the disease and calibration models in EPIC: “After exclusion of 22 432 prevalent cases of any cancer other than nonmelanoma skin cancer at the date of enrolment, 10 208 subjects in the top and bottom 1% of the distribution of the ratio of reported total energy intake (EI) to energy requirement (ER) (calculated from body weight) were excluded from the analysis to reduce the impact of outlier values.” **Bingham et al. 2002** ([PMID: 12484203](#)) already reported: “Data from individuals in the top and bottom 1% of the ratio of energy intake to estimated energy requirement (calculated from height and body weight) were excluded from the analysis.”

In datasets, the variable **EXCLEIER** is based on the EI/ER ratio. The bottom 1% of the distribution is coded as 1-Low, the top 1% of the distribution as 3-High, and the rest as 2-Medium. Data managers provide the full datasets, so please exclude subjects who have low (1) and high (3) EI/ER – this is generally mentioned in study exclusion flowcharts of manuscripts – and use data related to subjects who have EXCLEIER 2.

The *total energy intake* (kcal/d) is calculated as the sum of the energy intakes from the consumed foods.

The *estimated energy requirement* (EER) (kcal/d) is calculated based on “Energy and protein requirements the 1985 report of the 1981 Joint FAO/WHO/UNU Expert Consultation” (<https://doi.org/10.1111/J.1467-3010.1987.TB00040.X>).

For dietary questionnaires:

Males > 60 y:

$$\text{EER} = [8.8 * W + 1128 * H / 100 - 1071] * 1.55$$

Males 60 y > x > 30 y:

$$\text{EER} = [11.3 * W + 16 * H / 100 + 901] * 1.55$$

Females > 60 y:

$$\text{EER} = [9.2 * W + 637 * H / 100 - 302] * 1.55$$

Females 60 y > x > 30 y:

$$\text{EER} = [8.7 * W - 25 * H / 100 + 865] * 1.55$$

where W is weight in kg and H is height in cm.

12. Access to food lists of EPIC baseline dietary questionnaires

The *Mother matrix* is the simplified list of EPIC dietary items (foods, recipes, decomposed items, ingredients) classified according to the EPIC classification.

If some re-coding or analyses are to be done **at the food level**, a simplified matrix may be prepared.

Access to the whole or part of the *Mother matrix* can be requested from EPIC-Diet@iarc.who.int.

The presentation **2022_03_18_Mother-Daughter_Matrix_use.pdf** explains how FFQ food items were handled. It can be obtained upon request to epic-diet@iarc.who.int.

13. What are the intakes of food groups and subgroups and of nutrients in EPIC?

In 2009, descriptive analyses were run to provide intakes of EPIC food groups, selected items, and ENDB nutrients for both baseline dietary questionnaire data and 24-hour dietary recall data (available for the calibration subset in EPIC).

In the document [Food classifications](#), intakes are provided:

For baseline dietary questionnaire data: by sex, by country, by centre, N, mean, std, min, max, P1, P5, P10, P25, P50, P75, P90, P95, P99

In the sheet **DQ_Desc_FDGP**: by ECLASS food group and subgroup;

In the sheet **DQ_Desc_SEL_I**: by selected item.

For baseline calibration data: by sex, by country, N, mean, std, min, max

In the sheet **24_Desc_FDGP**: by ECLASS food group;

In the sheet **24_Desc_SEL_I**: by selected item.

In the document [Food components](#), intakes are provided:

For baseline dietary questionnaire data: by sex, by country, by centre, N, mean, std, min, max, P1, P5, P10, P25, P50, P75, P90, P95, P99

In the sheet **DQ_Desc_Nut**: by ENDB nutrient.

For baseline calibration data: by sex, by country, N, mean, std, min, max

In the sheet **24_Desc_Nut**: by ENDB nutrient.