Procedures for imputing component values in FCDBs

EuroFIR AISBL

Public version 2019

Original published as a confidential Deliverable for:
EuroFIR Network of Excellence (Grant Agreement No. 513944)
WP2.2 Composite foods, Deliverable D2.2.13, Version 5.6.2007

First published in Belgium in 2010 by EuroFIR NEXUS (Grant agreement ID: 265967)

40 Rue Washington

1050 Brussels

Belgium

www.eurofir.org

2019 – 01

EAN 9782930672014

9 782930 672014

CONTENTS:

1. Introduction

All component[1] or nutrient values of food items can hardly ever be analyzed for a national food composition databank (FCDB). For nutrient intake calculations, researchers and nutritionists hope for a complete dataset which eliminates missing nutrient values (Buzzard et al., 1992; Slimani et al., 2007). In principle, the compiled databank should be as complete as possible so that nutrient content of a diet is not underestimated. Therefore, the coverage in nutrient values in a FCDB may be different from the initial (reference) database (Becker et al., 2007, Greenfield and Southgate 2002). In the compiled database the gaps in nutrient values are covered by estimated nutrient values and thus the rules of estimation play an important role.

The management of a FCDB consists of two approaches (Westenbrink and Oseredczuk 2007, Feinberg *et al.*, 1991): First, storage and evaluation of collected composition data, primarily analysed data, from different sources constituting the initial database. Second, a selection of the initial database is aggregated for a selection of foods and components to an aggregated (compiled) database. From compiled database, datasets can be extracted for specific purposes (for example, a published Food Composition Table). In the management procedure the number of missing values is the starting point for quality evaluation of the user FCDB (Buzzard *et al.*, 1991).

The compiler receives original component values from chemical analyses which represent data obtained from the direct method, i.e. from original chemical analysis of local foods. Chemical analysis is also the most detailed level of data with strict determinations for each constituent, for an analytical method and for each food sample. In laboratories, researchers usually deduce statistical parameters with the results. Both the analytical details and statistical parameters that are available should be included with the results entered into the initial database of the FCDB. Chemical analyses may also express qualitative measures (e.g. trace, less than, more than, below detection limit).

For many nutrients or components, analytical results are only available for a few main food sources. Therefore, there are missing values in the initial database. The FCDB compiler aims to produce a complete compiled database of FCDB where the number of missing values is minimized. Component values can be estimated by indirect methods (by imputation or by borrowing data from other databases) in the compilation of a FCDB. The compiler may also apply imputation process to values from other databases and values from literature as well as to original analytical results of local foods.

Imputation is a compilation process to replace missing values with estimated values and includes a variety of procedures. Generally, imputation uses the existing data in the database for estimating a value. Mathematical operations, expert estimation and logical decisions are needed to prepare a usable compiled database with complete data sets for each food item and for as many nutrients as possible.

Definition of the concept ‘imputation’? To clarify the term, we present a citation from dictionary: to impute – to attribute = to regard as a characteristic of a thing (Merriam-Webster’s Dictionary). Imputation is a process for producing ‘best estimate’ values where no analytical value is available. This process includes calculation procedures, statistical modelling and logical decisions.

[1] Term component is preferred but term nutrient is used as synonym for a limited selection of components.

2. Objective

The EuroFIR Work Package 2.2 'Composite foods' aimed to standardize procedures for calculating nutrient values of composite foods. Food composition databases (FCDB) are used for many purposes, including converting food consumption into nutrient intakes. Calculation software is typically used for this purpose. Calculations are used in research, catering, food manufacturing, educational sciences and health services.

For nutrient intake calculations a complete, a compiled[2] database is required and involves construction of new food items (especially composite foods and their recipes). Food preparation steps affect the nutrient content of mixed dishes and composite foods in various ways, and the compilation of a FCDB is modelling these steps and processes. Also manufactured branded foods have treatments which food composition databank must model. Compilation processes must be able to handle food characteristics to find similarities of food items.

This report is oriented towards imputation based on component values, but the derivation of components values according to chemical relations of isomers and components are not covered by this report. For example, carbohydrates calculated by difference or energy calculated from energy nutrients, are not considered as imputation and guidelines for these calculations are presented by WP1.8.

The recipe calculation procedure is used to produce nutrient values for mixed dishes and also used to produce component values for missing values for other food items. Thus, recipe calculation procedure is one of the useful procedures in imputation. This presentation collects applications, procedures and rules used in imputation procedures of component values in the FCDB management. The aim of this report is also to elucidate the use of recipe calculations in the management of the FCDB and to propose recommendations for implementation.

[2] The stages of database (initial vs compiled) are discussed later in Chapter 5. See also EuroFIR Standards.

3. Imputation of missing values in the management of a FCDB

The starting point for the maintenance of a FCDB is to select and describe food items for the database. Let us now assume that a new set of analytical values for nutrients/components has been published by a governmental laboratory. The FCDB compiler assesses the validity and reliability of the results according to a standard operation procedure, e.g. applying proposed EuroFIR protocol (together with analysts). Acceptable values and documentation are entered into the initial FCDB. The database containing the analytical results with all relevant information on analytical conditions and literature sources is called an initial database. EuroFIR, WP1.3 Task Group 2, is developing standard operation procedures (SOP, Westenbrink and Oseredczuk 2007), which are closely related to national compilation guidelines (for published information see e.g. Food Standards Agency 2005).

Processes for aggregation of nutrient values will be established based on both the analytical information and literature data from other laboratories. The aim of imputation is to determine the most probable values or 'selected values' or 'best estimates' for components and food items where no analysed values are available (Charrondiere et al., 2002) in order to minimize the number of missing values in the compiled database.

Imputation may consist of several compilation cycles for construction of the aggregated database of FCDB (Sievert et al., 1989, Schakel et al., 1999, Slimani et al., 2007). The initial values and the aggregated data can be held within the same databank, but are most often managed in separate databases of the FCDB. For specific purposes a separate database may also be designed.

Imputation should be based on agreed common principles according to which the compiler searches for values for the missing components. Principles may be occasionally redesigned or adjusted in component-dependant decisions. When formulating these principles, similarities between foods and analytical characteristics of the components are considered. The compilers may also evaluate the scope and variations of international food composition database. The co-operation of compilers and analytical experts on common principles is useful. It is essential to discuss in interaction about various, alternative possibilities for data selection and imputation solutions. A large process of imputing missing values was carried out in the EPIC Study in ten European FCDBs (Slimani et al., 2000; Charrondiere et al., 2002; Slimani et al., 2007).

Imputation process includes:

- Common guidelines and standard operating procedures (SOP) for acceptable values
- Search and evaluation of literature values
- Calculation procedures based on analytical results and similarity of food items (imputation)
- Computing the new version of aggregated/compiled database

The compiler of a FCDB should be able to describe all the details and information used in imputation (Charrondiere et al. 2002; Burlingame 2004): All documentation, such as formulas, decision rules, figures, and food descriptions should be stored in the FCDB. The compilers must be aware of modifications in the FCDB verifying that intended changes occur and testing for unintended changes between versions of the FCDB (Sievert et al., 1989).

To achieve a complete recording and to store a repeatable process, imputation procedure should preferably be modelled in the database itself. It is recommended to formulate the structure of the FCDB in a way that the imputation procedures are part of the management of the FCDB. The recipe calculation procedure may be used for estimating missing nutrient values because it maintains the information of ingredients and information of the preparation processes of the composite food and its ingredients. The recommendation for recipe calculations is presented in the recent EuroFIR report (Reinivuo et al., 2007).

Some imputation procedures may be carried out manually outside of the FCDB. This is the case when labelling information for branded food products is collected and the means of the contributing nutrient values are calculated on separate data sheets (e.g. Excel spread-sheets etc.) without entering the individual data in the FCDB. In this case the documentation of imputation may be generated and stored outside the FCDB. Such a system has limitations because the definitions in the FCDB may change and out-dated information may exist on separate sheets (Unwin 2000). In these cases, updating may be difficult if the imputation process including calculation procedures is not documented appropriately in the manual records. Neither can the imputation be repeated in cases where there happens a change in definitions (e.g. coefficients of vitamin A).

In the EPIC Study (Slimani *et al.*, 2007) ten European countries formulated a common strategy for data documentation. A specific component-group strategy was used for imputation (partly published). Foods with high consumption were prioritised. The imputation itself was fulfilled firstly by searching best available alternatives from the same food in each national FCDB and secondly by deriving nutrient profiles from foods with similar nutrient composition. After imputation procedures the completeness of subsets[3] from national FCDBs improved remarkably ranging from 81% to 100%. After final evaluation and systematic imputation, a minimal proportion of missing values remained (under 1%).

Alternative solutions for imputation of nutrient values in a FCDB are discussed in detail below including:

- Analysed without any quantitative result (4.1)
- Borrowing or calculating of a value from a similar food (4.2., 4.3)
- Use of labelling data of industrial food products (4.4) and
- Expert decision (including use of logical zeros) (4.5).

[3] Subset of food items, requested by the EPIC Study, from composition matrix of national study centre.

4. Alternative imputation procedures for estimating component values

4.1. Transformation of quantitative results

Analytical results from a laboratory may contain qualitative results such as 'under detection limit' or 'traces' and they are thus recorded as such in the initial database. In the aggregated/compiled database, values may also be recorded as qualitative results and be published as such in a printed or on-line published food composition table. However, when preparing datasets for nutrient intake calculations (e.g. consumption survey or recipe calculation) the compiler should transform these qualitative results to quantitative values, i.e. choose either 0, half of the detection limit or some other value. Value type recommended by EuroFIR is separating definitions for each of these qualitative results (Becker et al., 2007). The recommendations in these circumstances available (Greenfield and Southgate 2003) are in favour of using half of the detection limit. Zero value for a nutrient is used when no content for this nutrient was analytically detected.

4.2. Imputing values from similar food(s)

The simplest method of imputation is to insert the analysed value of a food item for the other food item of the same origin. A value can be copied as such from a similar food item where an analysed value exists (Schakel et al., 1997). For example, for rowanberries nutrient values can be imputed from those of lingon-berry. The relationship between two or more food items is determined by contributing food items or inter-item link. This relationship should be documented in the FCDB, possibly as a recipe or as written text. The imputation procedure should be documented and stored for each value when procedure is separate and individual.

An average of existing analytical values in a group of similar food items can also be used as an estimated value for other food items in the group. For example, we have mandarin (*Citrus reticulate*) as a food item of citrus fruit with no analytical results for minerals or carotenoids. We can calculate the average mineral and carotenoid content of all the other citrus fruits and apply the average values for the mandarin. In the compilation software we can model this as an aggregated food (by recipe) including citrus fruits as ingredients.

Also, other statistical parameters like median of analytical results for similar food items can be used to replace a gap in nutrient values. In averaging, weighting can also be used for calculating representative values. The weighting factors can be derived from food consumption data of the population or from food market shares. For each value the background information should be stored whenever possible.

Averaged nutrient profile for a food group can be used to impute the nutrient content for a food item with missing values. For example, we have no analytical values available for α-carotene, β-carotene in chive (*Allium schoenoprasum*). Missing values could be estimated from the average content of α-carotene and β-carotene in leafy vegetables, as was done in the example of mandarin orange above. However, a more sophisticated formula can also be used including information of the nutrient profile of chive into the model and accepting only values of α-carotene and β-carotene in leafy vegetables with similar carbohydrate content.

If we have analytical values for carotenoids in full-fat milk, we can estimate carotenoids for other liquid milk products with different fat content by model according to the relation of fat contents. β -carotene for full-fat milk (fat 3.5%) is 20 µg /100 g, the formula for β -carotene content in semi-skimmed milk (fat 1.5%) is:

x = 1.5/3.5 * 20 = 0.43 * 20 = 8.57 µg

where the relationship of fat content in semi-skimmed and full-fat milk is used for the coefficient of the relationship. In this model it is assumed that β-carotene content in milk products varies according to fat content.

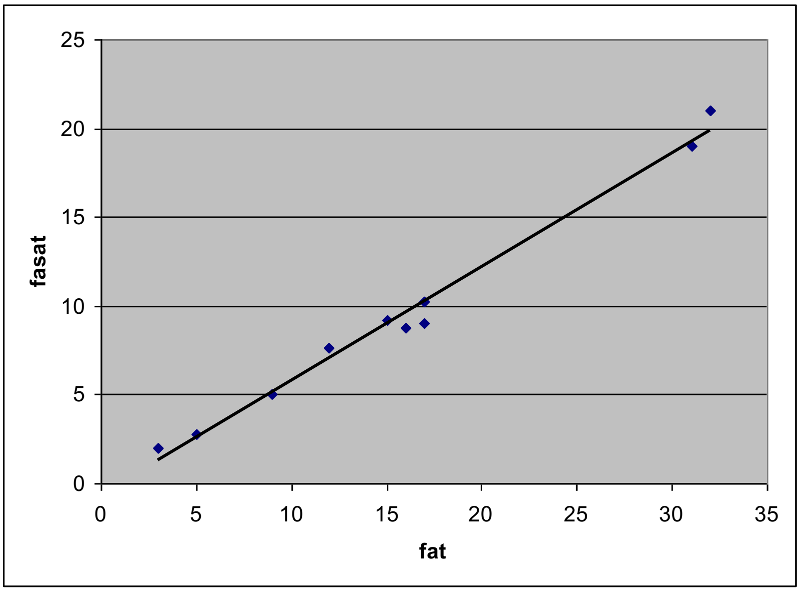

Figure 1. Example of a regression line between contents for total fat and saturated fatty acids in milk products

A common example of calculating from similar foods involves fatty acids. A regression coefficient can be calculated from a set of analytical data on fat and fatty acid content in, for example, milk products. It is assumed in this model that contents of a nutrient correlate with values of another nutrient (e.g. total fat and fatty acid profile). As an example of more sophisticated methods, missing values can be predicted by regression models to interpolate a value from the others. For instance, the fatty acid content of a cheese, with unknown content, can be extrapolated by regression between fat and fatty acid contents of other cheeses in the databank (Figure 1).

4.3. Calculating a value from an analysed value in a different form of the same food

If we have fibre content for unpeeled apple and for apple peels separately, we can calculate fibre content of an apple eaten with peel. The FCDB has the information about the proportion (13%) of peels in a whole apple. Thus, the proportion of edible matter is 87%.

For fibre content of peeled apple, we use the formula:

x = (a – 0.13*b) / 0.87

where a = fibre content (g) in whole, unpeeled apple (100 g),

0.13 = proportion of peel in the apple,

b = fibre content in peels (g/100g) and

0.87 = edible proportion of the peeled apple, to express results per 100g.

In this case the compiler probably calculates the required values outside the FCDB and enters the appropriate documentation according to which the calculations can be repeated.

In some cases, analytical results may be expressed per dry weight whereas a FCDB expresses nutrient values per fresh weight. For example, the sodium content per fresh weight (FW) of a food item can be calculated using water content and analysed sodium value per dry weight (DW) of the same food.

If water content is 80 g/ 100 g fresh weight, there is 20 g dry matter /100 g. The analysed value for sodium content is 200 mg Na/ 100 g (DW).

200 mg Na/ 100g (DW) = x mg Na/20 g (FW).

x = 20 g * 200 mg / 100 g = 40 mg Na/ 100 g (FW).

In some cases, the statistics of composition in food items of a food group or selected food items is the best estimate for a component value. Langual facets might be useful as selection criteria. As an example, in future the compiler can manage the labelling information of branded food products to model the most probable nutrient content as an average or as the average content weighted by consumption in various age groups.

The calculation of energy content, sodium from salt or carbohydrate content by difference follow basic chemical calculations and have guidelines in the manual of management of food composition data (Greenfield and Southgate 2003). The calculation of energy, sodium from salt, or carbohydrate content by difference follow basic, often chemical, calculations and have guidelines that can be found in literature (Becker et al., 2007, Greenfield and Southgate 2003).

4.4. Imputation of component values based on labelling information of food brands

For imputation of component values according to labelling information of manufactured food products it is possible to select from two main approaches which are presented here. The compiler can enter the main branded food products and their labelled nutrient values into the FCDB and use these in data aggregation or the compiler can enter industrial ingredients into the FCDB and construct a generic food entity with an averaged recipe in order to calculate component values.

A. Composition of generic food from branded food products

Using labelling information for calculating nutrient values is basically an example of recipe calculation. However, in the case of manufactured foods, the amounts of ingredients are missing: all that is known is the order (ranking, see Becker et al., 2007) of ingredients by weight, as indicated on the labels. Retro-calculation of the quantity of ingredients in an industrial recipe is typically done by a trial-and-error technique where a solution is searched until a calculated value sufficiently close to the known variation in values is achieved (Schakel et al., 1997). Applications can use labelled nutrient values of each branded food products where the limitation is that labelling concerns only a minor part of components of the databases.

As an example, The McCance and Widdowson's printed food composition table (Food Standards Agency 2002) contains nutrient values of tortilla chips. The printed table describes that values are produced from analysis of 20 samples representing 6 brand names of maize chips. In this case, the Food Standards Agency does not publish the brand names in the printed table. However, the data bank obviously has details of brands for documentation as seen elsewhere in the printed table. There has been a general principle not to publish brand name information in national food composition tables or only in limited situations. The Food Standards Agency (2007) of UK has formulated rules according to which it is judged to express names for branded food products in the analytical reports.

To enter manufacturers' data into the FCDB, labelled information for all branded food products with a significant share in the food markets will be collected in order to record both ingredients and nutrient values. Each brand can be a food entity of the FCDB with all the labelled nutrient content entered into the initial database of the FCDB (Unwin 2000). The determination of inter-item relationship or contributing foods is needed between each brand and an aggregated food item of the FCDB. This information can be recorded as a "recipe" using national consumption or possibly weighted by market shares.

The background of labelled nutrient values and their quality is not always known. The aggregated food product in the FCDB can have as ingredients the main brands, and so nutrient values can be calculated according to the recipe calculation procedure (Reinivuo et al. (2007). This is the case when the compiler enters all the labelled nutrient contents for the main branded foods into the initial database and then calculates representative component values for these brands. However, the nutrient content of the aggregated food can be only calculated according to stored labelled nutrient values and so remains limited in scope.

For labelled nutrient values, it is very important to evaluate, if fortification of the product with some nutrient is included in the labelled values. The separation between natural nutrients and synthetic isomers of the same nutrient should be separated when possible. Fortification may differ by country, and borrowing a value from a published Food Composition Table contains here a possibility of misinterpretations.

B. Composition of generic foods from average ingredients

Based on the lists of ingredients in the branded food products, the proportions of ingredients can be estimated by trial and error as above (retro-calculation), to create a recipe of an average food product. In this procedure only the ingredients are stored as food items in the FCDB. The most representative brand(s) can be chosen by market shares and its/their ingredients used for making a recipe for an average food product (Schakel 2001; Charrondiere et al., 2002).

Making a recipe for commercial branded food products necessitates a good coverage of industrial ingredients in the FCDB. Industrial ingredients may be modified for specific purposes in product development of particular types of food, they are not usually available for households, and so the composition of them may not be available or is not always clear (Ubbink & Krüger 2006). Among industrial ingredients there are novel foods and ingredients that are novel foods (Burlingame 2004). European FCDBs have only limited data for novel foods.

In fact, neither of these procedures is completely satisfactory. However, there hardly are resources to analyse all nutrients in all branded food products. Labelling information contains only the main nutrients, which do not cover the whole selection of components of the FCDBs.

A third approach would combine the first two methods, using label information for macro nutrients and recipe calculation to impute the values missing from the nutrient labelling.

4.5. Using logical and assumed zeros

The compilers usually make many searches of data in scientific literature and on websites of food industry companies. However, situations still remain where no data are available at all. In some cases a value may be imputed based on expert estimation and/or nutrient profile of a food group. Expert estimations may conclude to use zero values based on logical decisions or other type assumptions. These concepts can be separated by following definitions: Logical zero is based on specific components of a food item or a food group. Assumed zero is based on more general view of food characteristics. The separation of these concepts can be useful in making algorithms of imputation.

The decision to use logical zeros is based on either a component profile, such as zero value for protein in white sugar, zero for fibre in oils or origin of the food, such as vitamin B12 in foods from vegetable origin (Schakel et al., 1997). In the national data sets that were included for the EPIC study the number of missing values varied from 17% to 57%. The use of logical zeros enabled the task of data imputation to be focused on real missing values (Slimani et al., 2007).

As a conclusion, imputation procedures demand detailed information on components, on the food items analysed, on the food items available in the FCDB and on ingredients of recipes. For these reasons, a co-operation of compilers and analysts is essential in order to arrive at a reasonable solution. In the documentation of the FCDB, the procedures used for imputing or calculating nutrient values must be clearly described and documented.

5. Data stages of a FCDB and imputation process

5.1. Structuring a usable compiled food composition databank

Finally, we elucidate the levels of the FCDB to describe the views and differences in maintenance processes of composite foods.

The aim of a compiled database is to give a complete nutrient profile for each food item of the database. Compiled databases of a FCDB tend to have a wider selection of food items than an initial database with several variable ingredients and formulas of basic food items. There are compiled FCDBs that are relevant for a limited purpose like for a specified dietary survey or intervention. There are also compiled databases for the whole population and its professional experts. A compiled database should always be based on an initial database. The main selection criteria are the coverage of food items that are available for the diet of the population and the coverage of nutrient values under concern.

For intake calculations, for converting dietary data into nutrient intakes, the database is recommended to cover values for all basic nutrients (proximates) at least, or for all prioritized nutrients, components prioritized by EuroFIR (WP1.5) or for nutrients with special interest. The compiler or the final usable database of the FCDB may have to complete datasets for many nutrients before the gaps in interesting nutrients are reduced as much as possible. It is recommended to do compilation process as complete as possible by FCDB compiler to minimize tasks needed. This responsibility means that the final user does not have to check data that covers the prioritized nutrients of the user's study.

The process filling in missing values to complete the FCDB matrix is called the imputation of component values. Imputation means a process where all the analytical data of the matrix is used to estimate nutrient values in cases where no analytical results are available. The imputation is usually based on assumptions of a similar composition with a food item where analytical results are available.

Mathematical calculations, logical definition, statistical measurement and other justifiable procedures can be used in data aggregation. The selection of the best values usually demands comparison between values in literature, values from foreign databases, calculated values based on domestic analytical data, and labelling information of branded food products. Decisions must be made after comparisons and it is useful to combine the expertise of analysts and compilers in this task. Is the food for which we have a foreign analytical value comparable to the domestic food item or food samples? Is the method used in the published analyses the one that is preferred? Is the nutrient defined similarly and is the mode of expression (unit of measurement) the same in the analytical work as in the FCDB? The process of data aggregation/compilation usually demands several cycles before all the values of a nutrient are acceptable.

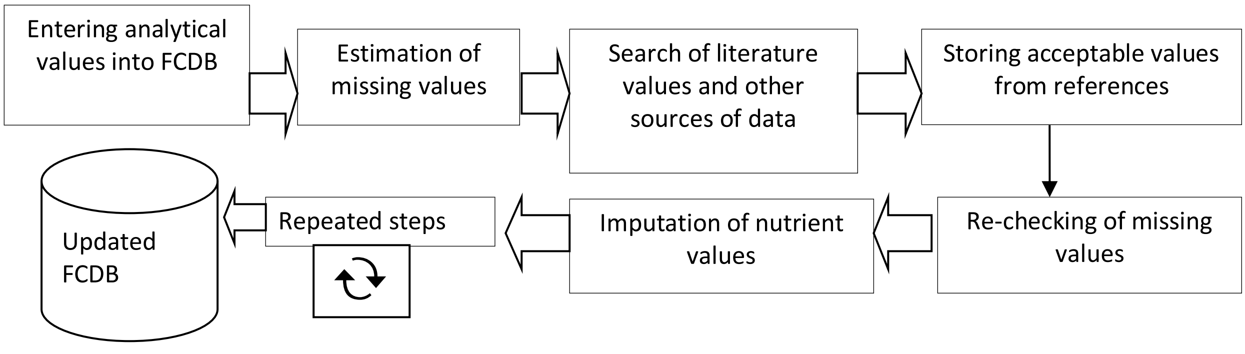

Figure 2. Possible flow chart of imputation processes for estimating and compensating missing values

5.2. Initial database

The analytical values on the composition in food items comprise the basis for a food composition database. The analytical values are documented concerning sampling, brand names (of individual or pooled samples), sample handling, analytical methodology, levels of recoveries, laboratory and publication. Documentation, quality evaluation and acceptance of new data must be carried out according standards when entering data into a FCDB (SOP, Westenbrink and Oseredczuk 2007). Incoming data can be quality assessed and described by statistical parameters such as numbers of samples, mean, median, minimum and maximum values, and standard deviation. These parameters can be calculated for individual component or by food groups and they are useful for comparing different versions of the database.

Initial data is also used to calculate aggregated composition by mathematical operations. An aggregated nutrient profile for a food group can be shown (Schakel et al., 1997) as a descriptor of the data but the nutrient profile can also be used as rough estimates for the composition of a new food item belonging to this food group. Component values can be imputed also by more sophisticated mathematical models that model the content information. As an advantage of modelled calculations is that they can be re-scrutinized after new initial data has been added to the FCDB.

The initial database can contain several different values for one nutrient of a food item (Becker et al., 2007). These values can describe nutrient composition of the food item by species, variety, seasons, sampling years etc. In principle, a mean or weighted mean is calculated from these single values into the user database. The initial database and the aggregated database can contain component values for food items that will not be included in the compiled database such as published food composition tables or extracted database for users. Such food items (e.g. potato peels or chicken skin) may be useful for the compiler of a FCDB but they are not usually recorded in dietary recalls. There can be non-analytical data in the Initial database. Data taken from a food composition table or from label information might not be analytical. The initial database contains data that have been entered into the FCDB and which are documented with reference code, date, etc. A short flow of management procedures is presented in Figure 3 (see more detailed presentation by Westenbrink and Oseredczuk 2007).

Considering the increasing proportions of manufactured food products in the European diets, the food industry is an important partner in information of composite foods (Slimani et al., 2007). Typically, the labelling data for industrial food products is collected more or less manually from packages, leaflets or best from companies' websites (Schakel 2001). Data transfer from food industry databanks would help the data collection needed for the maintenance of nutrient content and ingredient data of manufactured food products. Data availability from food industry is especially important concerning novel ingredients which are used in food industry. The status of European FCDBs revealed that almost no novel foods were available in European databases in 2005.

Validity of imputed values of a FCDB is determined as the ability of this procedure to approach the average value that probably is available in the food consumed by the population. The validity of imputed values has been tested with dietary data of the population (Charrondiere et al., 2002). It has been also suggested that it is possible to test validity of imputed values by feedback from on-line databases (the preponderance of imputed values in critical comments). Structuring of the FCDB needs qualification of the consistency by logical operations.

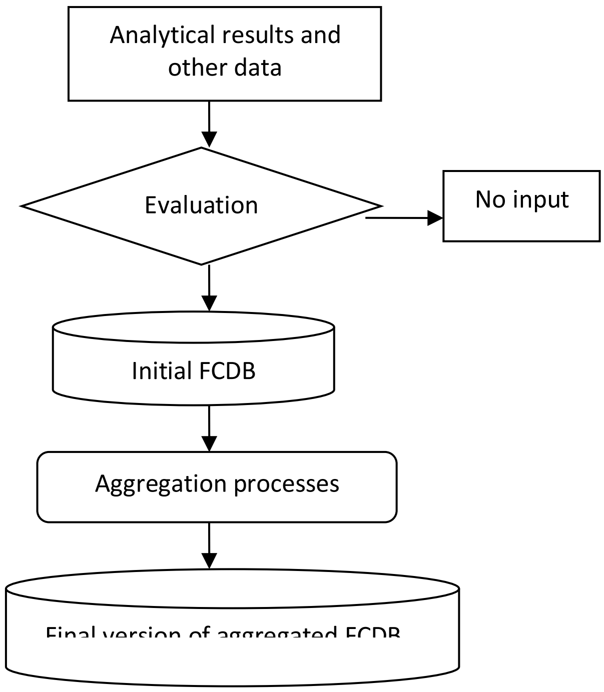

Figure 3. The main steps in the maintenance of a FCDB

5.3. Imputation procedures and further EuroFIR tasks

Separate guidelines for derivation of component values by using relationship of other components is needed (WP1.8). This includes derivation of energy, sum of fatty acids or fatty acid triglycerides, vitamin A RE or vitamin E TE. For component values of dietary supplements derivation of components content may be derived according to chemical composition (e.g. calcium from calcium phosphate).

The documentation of compositional values produced by imputation procedures is a part of value documentation in the EuroFIR data interchange standard. By standard coding the characteristics of nutrient values can be assessed similarly between FCDBs. This work needs to be fulfilled by WP1.8 in EuroFIR.

Recommendations: (1) General rules and guidelines for decisions on quality index should also be planned; (2) Co-operation with industrial partners is needed to apply an open data transfer from industrial data banks to compilers.

6. Recommendations for imputation in European FCDBs

Most printed food composition tables publish analysed nutrient values of raw or unprocessed foods. In use of food composition data, we need as complete a set of values as complete as possible and a limited number of missing values. Zero values are used in a FCDB only when absence of the nutrient is confirmed. Imputation is a common process in the FCDB management for the purpose of filling best estimates for missing values. The following recommendations concern imputation:

1. Imputation is a collection of estimation methods for producing component values that are considered as the most probable values or 'best estimates' for components where no analysed values are available.
2. Imputation of a component value is primarily based on the component value of a similar food in another form or a food with same origin.
3. Compilers and analysts should decide the common principles according to which the compiler searches applicable values for the missing values.
4. The loss of the component due to preparation (e.g. from fresh to cooked) and the need for retention factors have to be assessed with each imputed value and standard retention factors used (EuroFIR WP1.8).
5. Each imputed value must be categorized according to EuroFIR guidelines (WP1.8).
6. Each imputed value should have documentation of the background such as item-link, recipe, description, formula, or decision rule.
7. Quality index of each imputed component values should follow EuroFIR guidelines. This is the confidence code for which a determination method still needs to be developed.
8. Decisions for use of logical or assumed zeros and trace values should be made before starting imputation in order to decrease the workload.
9. Co-operation of compilers and analysts is needed to construct decision rules for value imputation and decisions on qualitative analytical data.
10. For prepared or mixed dishes the estimation of missing values is based on the use of local recipes from local (national) cookbooks and this data is entered into the FCDB.
11. The nutrient contents of prepared and mixed dishes are estimated by using local (national) coefficients for yield of water and fat.
12. Ingredient lists from labels of branded foods can be used for imputation as an average recipe of an aggregated food or as individual contributing brand items of an aggregated food.
13. Checking the quality of imputed values in a published version of FCDB (food composition table) is advisable by verifying with double decisions a random selection of imputed values. In the case of enlarged analytical resources, component values for a selection of food items with imputed values can be analysed and a comparison study between analysed and imputed values carried out.
14. The compilers must evaluate the repeatability of imputation and search unintended changes between versions of the database by food group statistics.
15. All the recommended checks may be included in the automated updating program to carry out a complete checking protocol.

7. References

Becker W., Unwin I., Ireland J., Møller A. Proposal for structure and detail of a EuroFIR standard on food composition data. I: Description of the standard. http://bit.ly/2Ge5RhO (accessed 31.01.2019)

Becker W., Unwin I., Ireland J., Møller A. Proposal for structure and detail of a EuroFIR standard on food composition data. II: Technical Annex. EuroFIR AISBL http://bit.ly/2swvDWy (accessed 31.01.2019)

Burlingame B. Fostering quality data in food composition databases: visions for the future. JFCA 2004;17: 251-258. https://doi.org/10.1016/j.jfca.2004.03.026 (accessed 13.02.2019)

Buzzard M, Price K, Feskanich D. Preparation for dietary data analysis: how the structure of the database influences data collection and analysis, and vice versa. In 'The diet history method' (Ed. L Kohlmeier), pp 39-51. Smith-Gordon and Company Ltd, London 1991. (ISBN13 9781854630667, 13.02.2019)

Charrondiere R, Vignat J, Møller A, *et al.* The European nutrient database (ENDB) for nutritional epidemiology. J Food Composition Anal 2002;15:435-451. doi:10.1006/jfca.2002.1089 (accessed 13.02.2019)

Feinberg M, Ireland J, Favier J-C. Validated data banks on food composition: concepts for modelling information. World Review of Nutrition and Dietetics 1992;68:49-93. doi.org/10.1159/000420940 (accessed 13.02.2019)

Food Standards Agency. The McCance and Widdowson's The composition of Foods. 6th ed. Cambridge 2002, Royal Society of Chemistry. (7th Edition, http://dx.doi.org/10.1039/9781849735551, accessed 13.02.2019)

Food Standards Agency. Guidelines for undertaking surveys, 2005. http://bit.ly/2BBjEMi (accessed 13.002.2019)

Greenfield H, Southgate D. Food composition data, Production, management and use. FAO, Rome, 2003. http://bit.ly/2Wws3JX (accessed 31.01.2019)

Reinivuo H, Laitinen K. Proposal for the harmonization of recipe calculation procedures. Draft April 2007. (Available later as http://bit.ly/2SysGUU, accessed 13.02.2019).

Schakel S, Buzzard M, Gehardt S. Procedures for estimating nutrient values for food composition databases. J Food Compos Anal 1997;10:102-114. http://bit.ly/2GGqdAG (access 13.02.2019)

Schakel S, Harnack L, Wold C, van Heel N, Himes J. Incorporation of trans-fatty acids into a comprehensive nutrient database. J Food Compos Anal 1999;12:323-331. https://doi.org/10.1006/jfca.1999.0836 (accessed 13.02.2019)

Schakel S. Maintaining a nutrient database in a changing marketplace: keeping pace with changing food products – a research perspective. J Food Compos Anal 2001;14:315-322. https://doi.org/10.1006/jfca.2001.0992 (access 13.02.3019)

Sievert Y, Schakel S, Buzzard M. Maintenance of a nutrient database for clinical trials. Controlled Clinical Trials 1989;10:416-425. https://doi.org/10.1016/0197-2456(89)90006-8 (access 13.02.3019)

Slimani N, Charrondìere R, van Staveren W, Riboli E. Standardization of FCDBs for the European prospective investigation into cancer and nutrition (EPIC): General theoretical concept. J Food Compos Anal 2000;13:567-584. https://doi.org/10.1006/jfca.2000.0910 (access 13.02.3019)

Slimani N, Deharveng G, Unwin I, *et al.* The EPIC nutrient database project (ENDB): a first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. Eur J Clin Nutr 2007; Sep;61(9):1037-56. Epub 2007 Mar 21. DOI: 10.1038/sj.ejcn.1602679 (access 13.02.2019)

Ubbink J, Krüger J. Physical approaches for the delivery of active ingredients in foods. Trends in Food science and technology 2006;17:244-54. https://doi.org/10.1016/j.tifs.2006.01.007 (accessed 13.02.2019)

Unwin I. EUROFOODS guidelines for recipe information management. J Food Compos Anal 2000;13:745-754. https://doi.org/10.1006/jfca.2000.0911 (accessed 13.02.2019)

Westenbrink S, Oseredczuk M. Food composition databases: The EuroFIR approach to develop tools to assure the quality of the data compilation process. Food Chemistry (2009);113:759–767. doi:10.1016/j.foodchem.2008.05.112 (accessed 13.02.2019)