FOOD COMPOSITION DATA QUALITY: ASSESSING VARIATION AND QUALITY OF DATA FROM DIFFERENT SOURCES

Mark Roe [1], Jenny Plumb[1], Susanne Westenbrink[2], Luisa Oliveira[3], Isabel Castanheira[3], Marine Oseredczuk[4], Aida Turrini[5] and Paul Finglas[1,6]

[1] Institute of Food Research, Norwich, UK, mark.roe@ifr.ac.uk & paul.finglas@ifr.ac.uk
[2] National Institute for Public Health and the Environment (RIVM), Bilthoven, NL; susanne.westenbrink@rivm.nl
[3] Coordenadora da Unidade de Observação e Vigilância, Departamento de Alimentação e Nutrição, Lisbon, PT, Luisa.Oliveira@insa.min-saude.pt & Isabel.Castanheira@insa.min-saude.pt
[4] French Agency for Food, Environmental and Occupational Health Safety (ANSES), Paris, FR; marine.oseredczuk@anses.fr
[5] Council for Agricultural Research and Economics - Research Center for Food and Nutrition, Rome, IT; aida.turrini@crea.gov.it
[6] EuroFIR AISBL, Brussels, BE; pf@eurofir.org

Abstract − Users of food composition data assume that all data is good quality and fit for purpose but quality can vary and depend on a wide range of factors. Different types of data, e.g. nutrients and bioactive compounds and data derived from different sources can differ in quality. The EuroFIR system for assessing data quality allows sources of variation to be evaluated. Evaluation of food description, component identification, sampling, number of analytical samples, sample handling, analytical method and analytical quality control allows users to assess data quality.

Keywords: Food composition, Data quality, Bioactive components, Food labelling

1. INTRODUCTION

Food composition data is an essential component of most investigations into links between diet and health. Data is used for a wide range of purposes including: research, public health nutrition and education, policy development, trade, food product development and clinical nutrition. Developments in information technology systems, particularly the move towards web based technology, have led to increasing access to food composition data. Recent changes to European Food Information Regulations mean that all pre-packaged food producers will need to provide composition data for key nutrients and food composition data is increasingly available directly to consumers.

Most users take published values at face value and will not be aware of variation that may be associated with the values. Generic values that are intended to represent foods at a population level are published in most countries. However, even though they are intended to be representative of the foods described, individual samples may differ considerably because of natural or extrinsic variations and also because of apparent differences caused by analytical variations. Sources of natural variation in foods include season, growing conditions and post-harvest storage in plants and feeding regime and age in animals. Nutrients that are associated with structure and metabolic function tend to be less variable than nutrients that accumulate in particular parts of a plant or animal or nutrients that are unstable. Variation may also be introduced during processing or food manufacturing, e.g. changes in recipes, variation in amounts or types of ingredients.

Food composition data quality has been a feature of many European and International collaborative projects and formed the basis of the authoritative guide to the production, management and use of food composition data [1]. Quality approaches have been implemented and further developed by EuroFIR (European Food Information Resource) (www.eurofir.org) and INFOODS (International Network of Food Data Systems) (http://www.fao.org/infoods/infoods/en/). The EuroFIR quality framework [2] includes:

- Guidelines for the assessment of methods of analysis;
- A flow chart describing the food composition compilation process
- Development of a CEN standard 'Food data – structure and interchange format (EN 16104:2012)' [3];

- Thesauri for food description, component identification, units, matrix unit, acquisition type, method type, value type and reference;
- A recommended approach to calculation of composition based on recipe ingredients;
- A system for evaluation of quality of data from scientific literature or laboratory analytical reports;
- Peer review of data compiler organisations based on an ISO 9001 approach.

INFOODS also actively encourages improvement in quality of data by: providing international guidelines and standards (e.g. on food and component nomenclature; data interchange; compilation; data checks; conversions; sampling; data quality), assisting countries to compile and publish food composition databases and tables and by providing training.

The implementation of standards has enabled publication of harmonised data in national food composition databases and web technology has enabled publication of meta-data to support published values. However, many users are not aware of quality considerations so will use published values without any regard for variability, limitations of use or source of data. Published data is no longer limited to data from national composition datasets and scientific literature and nutrient composition of branded products is routinely available via manufacturer or retailer websites. Data on bioactive components in foods is widely used for research purposes and is also beginning to be made available to consumers through a variety of media.

These different types of data generally have the same quality issues as nutrient data but certain aspects, e.g. analytical methods used for analysis of bioactive components, may have increased impact on data quality. To enable users to make an informed decision as to the limitations of data, it is essential to make available documentation that describes the sources and methods used to obtain data. This paper describes quality considerations related to different types of food composition data and different aspects of the data.

2. DATA QUALITY CRITERIA

The EuroFIR system for evaluation of quality of data from scientific publications [4], including scientific literature and laboratory analytical reports includes six main criteria and was developed based on existing European systems. It is also compatible with the well-established USDA quality system [5]. The importance of each quality criteria will depend on the type of data and what it is intended to be used for. The system allows users to assess data quality based on common guidelines and a quality score can be attributed to each quality criteria. An overall quality score can be calculated but scrutiny of each quality criteria allows the user to consider the factors that are most important for their intended use.

2.1. Food description

Correct identification of a food is vital to ensure that food composition data is fit for purpose. Identification is usually based only on the food name provided but there are many factors that can influence nutrient content of a food and they should be included in the food name. For example variations may depend on cultivar, cut of meat, cooking and processing factors including addition of ingredients, fortification, removal or reduction of nutrients (e.g. fat, salt or sugar) and preservation methods. Food name should be unambiguous so that users can select the data that is most appropriate for their need.

Language can be a problem because data is often translated into English. It is not always possible to make a direct translation that is accurate and cultural differences can lead to different interpretations of meaning. Even within the same language there will be instances where the same food name can describe different foods e.g. prawn/shrimp, swede/turnip, or where the name is misleading, e.g. mincemeat does not contain meat and Bombay duck is a fish.

Facetted systems have been developed to describe foods and to distinguish between different foods that may not be fully described by the food name alone. The LanguaL (Langua alimentaria) system was developed by the United States Food and Drug Administration and has been modified and adopted for use in European countries as part of the EuroFIR initiative to better standardise approaches for food description in Europe. The

European Food Safety Authority has also developed a similar system of facet descriptors for use with the FoodEX2 food list that is used as the basis for dietary intake, exposure and risk assessments.

2.2. Component identification

Correct identification of food components is usually straight forward but can be complicated in components that are method dependent. For example dietary fibre is not a single food component but is a combination of compounds that are measured by methods intended to capture carbohydrates that are not digested. INFOODS publishes a thesaurus of component names that also identify the analytical method. EuroFIR also publishes a thesaurus of component names that is closely matched to the INFOODS thesaurus but does not include analytical method. Identification of the correct unit and matrix unit (e.g. per 100g edible portion or per kg dry weight) is also essential for correct use of food composition data. Some component values are derived by calculation from other components, sometimes with the use of calculation factors (e.g. protein calculated from total nitrogen using a nitrogen conversion factor). The calculation and conversion factors used should be clear and unambiguous.

2.3. Sampling

Sample collection is an important factor in food composition data to ensure that the value represents the food for which data is needed. Samples collected for national composition datasets are usually intended to represent foods consumed by the national population but data from other sources may simply be a snapshot of a food item rather than providing an average generic value. For some foods, factors such as season, geographic origin or brand may influence composition.

2.4. Number of analytical samples

The number of samples of a food that were analysed to produce a value gives an indication of how representative a value is. Ideally, statistical information about a value, such as standard deviation or range, should be provided to give an idea of variation.

2.5. Sample handling

Some components in some foods are likely to degrade following sample collection so it is important to ensure that samples have been handled correctly prior to analysis. Appropriate sample handling could include storage in inert containers, cooling or freezing samples and minimising storage time or exposure to oxygen. Prevention of contamination and complete homogenisation are also important.

2.6. Analytical method and quality control

Analytical methods may impact on values produced and an appropriate method should be used, wherever possible by a laboratory that is accredited for the method. Laboratory performance is also an important consideration and use of reference materials and/or proficiency testing improves confidence in published data.

3. DATA SOURCES AND TYPES

Food composition datasets are likely to contain data from a variety of sources and values that are produced in different ways. Ideally all values should be from samples analysed specifically for the dataset, but that is not usually possible so values may also be taken from scientific literature or other composition datasets. Values may also be calculated from other components in the same food or calculated based on recipe ingredients. Since composition data is available on many branded food products, data is often taken directly from manufacturers, in which case there is usually no information about the source and production of the data. Wherever possible, the source of the data should be provided to enable evaluation of data quality.

3.1. Compiled nutrient data

Nationally representative nutrient data is published in most countries and supporting information is often available. EuroFIR has encouraged documentation of all nationally published data and harmonised data from 28 countries is available through the EuroFIR FoodEXplorer tool (http://eurofir.org). Many European datasets can provide supporting information to allow evaluation of data quality and in some cases the original laboratory reports or scientific publications are available with the data. Although national composition data is carefully compiled it is not possible to update or validate all published values on a regular basis so in some cases

published values may be old. Where the date of publication or validation is provided it should be considered in data quality assessment because, even if the data was high quality when originally published, it may not necessarily reflect the composition of foods currently consumed.

3.2. Branded food data

Composition data for pre-packaged food products should be provided on product labels and is often also available from manufacturers or retailers websites. The data should be applicable to the product as sold but often there is not any additional information about the data. It is usually assumed that the data would have been analysed directly but food labelling regulations allow values to be calculated from ingredients or to use generic values from authoritative national datasets. Label data will not include information for nutrients that are not required by labelling regulations. Therefore if the data is to be included in datasets that include other nutrients, it is necessary to estimate the missing values from other sources. The approach is usually to map products to the food that is the best match in a national composition dataset however care must be taken to ensure that the food match is appropriate and in some cases further calculation to improve the estimation may be needed.

3.3. Bioactive component data

Data for bioactive components in food is almost all derived from direct analysis and is available either in specialist datasets or from scientific literature. There are several online databases of bioactives, including eBASIS [7], Phenol-Explorer [8] and the USDA Database for the Flavonoid Content of Foods [9]. Different examples of bioactive classes are contained in these datasets and all use quality systems to extract data from peer reviewed scientific publications. Quality assessment includes analytical methods used but methods for bioactive components are not yet as standardised as methods for nutrient analysis and proficiency testing schemes are not available.

4. CONCLUSIONS

Users of food composition data assume that all data is good quality and fit for purpose but quality can vary and depend on a wide range of factors.

Compilers of national food composition datasets take care to produce high quality data wherever possible but the quality of data from other sources may be less clear. Documentation of data sources and methods of production allows assessment of the quality to be made and users can judge whether or not data is fit for purpose. Standards for data production and publication are essential and help to ensure that data published from different sources is comparable. Training for producers, compilers and users of data is also important.

REFERENCES

[1] H. Greenfield and D.A.T. Southgate, *Food Composition Data: Production, Management and Use, 2nd Edition*, Elsevier, London and New York, 2003

[2] S. Westenbrink, M. Roe, M. Oseredczuk, I Castanheira, P. Finglas, EuroFIR quality approach for managing food composition data; where are we in 2014?, *Food Chemistry*, Vol. 193, pp. 69-74, 2016.

[3] CEN, *European standard. Food data – structure and interchange format. EN 16104:2012*, Available at <http://www.sis.se/en>. [Retrieved 02.06.2015].

[4] S. Salvini, M. Oseredczuk, M. Roe, A. Møller, Guidelines for Quality Index attribution to original data from scientific literature or reports for EuroFIR data interchange, EuroFIR, Available at <http://http://eurofir.org/>. [Retrieved 02.06.2016]

[5] J.M. Holden, S. A Bhagwat and K.Y. Patterson, Development of a Multi-nutrient Data Quality Evaluation System, *Journal of Food Composition and Analysis*, Vol. 15, pp.339-348, 2002

[6] Bioactive Substances in Food Information System. www.http://ebasis.eurofir.org

[7] JA Rothwell *et al*, Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content. 2013 http://phenol-explorer.eu/

[8] USDA Database for the Flavonoid Content of Selected Foods, Release 3.2 (September 2015). http://www.ars.usda.gov/SP2UserFiles/Place/80400525/Data/Flav/Flav3.2.