EuroFIR

European Food Information Resource

REPORT COMPARING USDA DATA QUALITY ASSESSMENT SYSTEM(S) TO EXISTING EUROPEAN SYSTEMS

Prepared by
Simonetta Salvini, Marine Oseredczuk, Mark Roe, Anders Møller & Jayne Ireland

EuroFIR TECHNICAL REPORT

EUROFIR WP1.3 TG4

REPORT COMPARING USDA DATA QUALITY ASSESSMENT SYSTEM(S) TO EXISTING EUROPEAN SYSTEMS

REPORT PREPARED BY WORK PACKAGE 1.3, TASK GROUP 4 AND WORK PACKAGE 1.8:

Simonetta Salvini, CSPO, Partner 20
Marine Oseredczuk, AFSSA, Partner 10

In collaboration with

Mark Roe, IFR, Partner 1
Anders Møller, DFI, Partner 46
Jayne Ireland, AFSSA, Partner 10

***ABSTRACT**:*

This document compares the USDA system, existing European systems and the proposed EuroFIR quality assessment system for evaluation of data from scientific literature and laboratory reports. The USDA system is designed for limited use, for specific components and within a single country, while the EuroFIR system aims to be more generic and is designed for application to a wide range of foods and components. The USDA system does not assess food description or component identification but five categories related to sampling and analysis are common to both systems. However the criteria assessed within the common categories are different so the scores provided by the two systems are not directly comparable although they do indicate quality of each category.

INTRODUCTION AND AIM OF THE DOCUMENT

Only a few systems exist to support quality assessment of food composition data. In order to create a validated EuroFIR databank system, it is important that all databases within the EuroFIR network evaluate the quality of food composition data according to the same criteria.

The data evaluation and the subsequent storing of decisions made according to such criteria should be sufficient to describe the quality of the single datum for each quality criterion.

According to Greenfield and Southgate (*Greenfield & Southgate, 2003*), the quality index of each original datum depends on **general evaluation categories or criteria** such as the description of the food, the component analysed, the analytical method, the sampling protocol, etc.… For each general evaluation criterion, **specific evaluation rules** can be determined, accompanied with a weighting or a grading scale for the determination of quality indices. The categories (or criteria), the rules and their weighting can depend on the type of data source (scientific literature, recipes, labelling, manufacturers data, old/archival data).

In addition to the common quality assessment of data that may be interchanged in EuroFIR, data quality assessment performed by European compilers should be comparable to evaluation by USDA compilers.

This document aims to compare the USDA system, existing European systems and the proposed EuroFIR quality assessment system for scientific literature and laboratory reports (*Oseredczuk et al., 2008*). This comparison is in part derived from a EuroFIR internal working document (*Oseredczuk et al., 2006*).

2. THE USDA QUALITY ASSESSMENT SYSTEM

During the last decades, the USDA has developed a system for the assessment of the quality of data to be included in their food composition database: the system has been adapted according to specifications characterizing each of the nutrients considered.

The first documents describing the quality assessment system were developed for the evaluation of the quality of analytical data for iron, selenium and carotenoids (*Mangels et al., 1993*; *Holden et al., 1999*).

These procedures were modified and expanded for multi-component evaluation (*Holden et al., 2002*), for the evaluation of flavonoids data (*Holden et al., 2005*), vitamin B2 and Vitamin K (*USDA system: Evaluation_Riboflavin.xls; USDA system: Evaluation_Vit K.xls)*.

Through the years the system has been based on the same five evaluation categories (Sampling plan, Number of samples, Sample handling, Analytical method, Analytical quality control) but the evaluation questions have been modified to be more objective. As a consequence, the rating scale of each of the five categories has also changed, moving from discrete steps to a continuous scale.

The rating of the five categories is summarized in a Quality Index (QI) (the sum of the five scores) and further categorized in the Confidence Code (CC), that is disseminated with the data, to indicate the level of confidence in the quality of each specific datum (*Holden et al., 2002*).

The USDA system has been developed by teams of experts, specifically nominated according to their area of specialty. The evaluation, at least in the latest versions, is carried out on a database management system that takes the evaluator through an evaluation tree, and automatically assigns points to each of the choices made (*Holden, 2006*).

In the following chapters, we will focus on the system used for the evaluation of Vitamin B_2 data, by HPLC.

An example: the USDA system for Evaluation of Riboflavin (Vitamin B_2) Methodology by HPLC with fluorescence detection

The system for the evaluation of Riboflavin (Vitamin B_2) is specifically designed to assess values derived from HPLC with fluorescence detection. The evaluation is divided in two parts, where the first is referred to as “evaluation” and the second as “method scoring” (see Appendix 1). The rating of each category is performed by answering a set of specific questions, here called criteria. According to the answer provided (pre-defined by the system) each criterion is assigned a score, which is then summed to obtain the total score for the category.

EVALUATION CATEGORIES

Five categories are scored separately, and the sum of the five scores provides the QI of the specific datum. The categories are reported below, with the original numbering of each criterion.

Sampling plan

The rating of sampling plans reflects the representativeness of the food sample to the relevant factor (e.g. brand, cultivar, market share, etc). Criteria to be evaluated and possible answers follow:

1. Was the sampling plan developed statistically?
   YES / NO / UNKNOWN
2. How many regions (countries) were sampled?
   1 / 2 / 3 / 4 OR MORE/ UNKNOWN
3. How many cities in each region or country on average?
   1 / 2 / 3 OR MORE/ UNKNOWN
4. How many locations in each city on average?
   1 / 2 OR MORE / UNKNOWN
5. How many different lots (or individual samples) collected at each location?
   1 / 2 OR MORE / UNKNOWN
6. Was the sampling done during more than one season?
   YES / NO / UNKNOWN

The scoring ranges from 0 to 20.

Sample handling

Proper handling of the sample units and composites is critical. The following criteria are questioned:

7. Is homogenisation necessary for this type of sample? (If "No skip to Question 11)
8. Was the sample homogenized?
   YES / NO / UNKNOWN
9. Was the homogeneity of the sample verified?
   YES / NO / UNKNOWN
10. Was information given on the equipment used to homogenize?
    YES / NO / UNKNOWN
11. Was only the edible portion used for analysis? (inedible portion such as seeds removed)
    YES / NO / UNKNOWN
12. Was moisture information given?
    YES / NO / UNKNOWN
13. Were the samples stored properly? (e.g. Frozen/ refrigerated?)
    YES / NO / UNKNOWN

The scoring ranges from 0 to 20.

Number of samples
The analysis of a small number of independent samples limits the ability to estimate the mean and variability (more details in Holden et al, 2002).

14. How many samples were analysed?
(The rating for number of samples is determined by the number of sample aliquots analysed independently. Repeated analyses of the same homogenate or the same composite validate homogeneity of the sample, but the number of analytical samples is one.) For mean values, use the count of individual, independent samples that went into that mean.
Possible responses: 1 / 2 / ….>12

Analytical Quality Control
This refers to the accuracy and precision in the day-to-day performance of the analytical method. Quality control material (QC) should be used with each batch on each day of analysis.

15. Was a control or reference QC material analysed with the analytical samples? (If "No", skip remainder of questions but if "Yes" continue with Question 16)
YES / NO / UNKNOWN
16. If a control or reference QC material was analysed what type was it?
Commercial QC material (SRM/CRM) / in-house QC material / Unknown or N/A
An in-house QC material is a material developed by the laboratory and used as a reference material.
17. If commercial QC material (SRM/CRM), how was the nutrient value listed?
Certified / Information or Reference / Unknown
18. How close were the QC material results to the expected values?
Within the expected range / Close to the expected range / Outside range or no information given
19. How frequently were QC materials analysed?
With every batch / Several times a day / daily / Occasionally/Never/Unknown
20. What was the coefficient of variation (%rsd) for the QC material?
<=5% / <=10% / <=15% / <=20% / >20% / Not given

The scoring ranges from 0 to 20.

Total Method Rating
This score is composed by three sub-scores (here indicated by a, b and c), related to sample processing, analyses, and laboratory validation of the method. The three sub-scores are summed to obtain the total method rating. In the USDA system these criteria are method specific. In the present case only analytical data for vitamin B2 obtained by HPLC with fluorescence detection are considered appropriate for evaluation.

a. Sample Processing
1. Was sample processing done under yellow light with low actinic glassware?
YES / NO / UNKNOWN
2. Was the pH maintained between 5.0 and 7.0?
YES / NO / UNKNOWN
3. Were standards dissolved completely with gentle heating?
YES / NO / UNKNOWN
4. If high starch foods analysed, were they treated with appropriate enzymatic digestion?
YES / NO / UNKNOWN / N/A
5. If high protein foods analysed, were they treated with appropriate enzymatic digestion or protein precipitation?
YES / NO / UNKNOWN / N/A

6. Was purity and efficiency of enzyme preparations tested?
YES / NO / UNKNOWN
7. Were samples kept frozen until beginning extraction?
YES / NO / UNKNOWN
8. Was a two stage extraction done (methanol+methylene chloride, buffer to pH5.5 in 2nd stage)?
YES / NO / UNKNOWN

The scoring ranges from 0 to 4, but is then summed up with the following analytical questions (b).

b. Analysis

9. Was the sample protected from light during analysis?
YES / NO / UNKNOWN
10. Was extract filtered before injection into the HPLC?
YES / NO / UNKNOWN
11. Were excitation and emission wavelengths reported for the fluorometric detection?
YES / NO / UNKNOWN
12. Was the purity of the standards checked?
YES / NO / UNKNOWN
13. Were standards prepared daily?
YES / NO / UNKNOWN
14. Was more than one concentration of standard used for external calibration?
YES / NO / UNKNOWN
15. Were at least 3 concentrations of a standard used for external calibration or was an internal standard used?
YES / NO / UNKNOWN

The scoring ranges from 0 to 4. The rating is summed to the score obtained for the sample processing questions (a) to obtain an interim score for method, ranging from 0 to 8.

c. Laboratory Validation of Method

1. Was a commercial reference material (CRM/SRM) analysed? (If "No" or "Unknown, skip to Question 4)
YES / NO / UNKNOWN
2. How is the nutrient value listed for the commercial reference material?
Certified / Information or Reference / Unknown
3. How close were the reference material results to the expected values?
Within the expected range / Close to the expected range / Outside range or no information given
4. What was the coefficient of variation (%rsd) for repeated analysis of the same material?
<=10% / <=15% / <=20% / >20% / Not given
5. What was the % recovery of the nutrient?
Not done / 95%-105% / 90-94% or 106-110% / 85-89% or 111-115% / 80-84% or 116-120% / <80% or >120% / Not given
6. Method results compared to separate independent method or separate laboratory - difference between results
<=10% / <=15% / <=20% / >20% / Done but results not given / Not given

The scoring ranges from 0 to 12. This rating is summed to the interim score for method, and the final Total Method Rating is thus obtained, ranging from 0 to 20.

QUALITY INDEX AND CONFIDENCE CODE

A QI is finally computed as the sum of the scores to the five categories.

In addition, also a CC can be assigned (A,B,C and D, or A, B and C).

Summary of scoring categories

SAMPLING PLAN
Scoring: some criteria have more weight than others
Possible scores: 0 - 20

SAMPLE HANDLING
Scoring: some criteria have more weight than others
Possible scores: 0 - 20

NUMBER OF SAMPLES
Scoring: some criteria have more weight than others, depends on the number of samples
Possible scores: 0 to 20

ANALYTICAL METHOD (for Vitamin B2 only HPLC is considered acceptable)
Scoring: a combination of sample processing, analyses and laboratory validation: some criteria have more weight than others
Possible scores: 0 - 20

ANALYTICAL QUALITY CONTROL
Scoring: some criteria have more weight than others
Possible scores:0 - 20

QUALITY INDEX:
Scoring: obtained as a sum of the five individual scores
Possible scores: 0 - 100

3. COMPARISON OF USDA VS EUROPEAN QUALITY ASSESSMENT SYSTEMS

Three Quality assessment systems in use in Europe have been identified. These are systems used, to various extents, by AFSSA - France, by CSPO - Italy, and by BASIS (see Appendix 2, 3 and 4). Recently, the BLS group in Germany has developed a quality assessment system, derived from the latest versions of the USDA system, but it is not commented on in this document.

Component identification

- AFSSA is currently not using component identification among its data quality assessment criteria (whereas the two other systems do), because AFSSA gets round that issue by using different codes for the different definitions of nutrients and assigning data to these codes. Nevertheless, even for assigning data to component codes corresponding to different nutrient definitions, compilers should first be able to distinguish between nutrient definitions and then identify precisely the nutrients to which the original data refer.

- BASIS is the only system proposing a relatively precise rating scale for component identification. As BASIS concerns only bioactive compounds, it is more conceivable for experts to define the analytical methods that give unequivocal definitions for only limited types of compounds. For example, NMR is judged by BASIS to be an unequivocal method of analysis for the identification of bioactive compounds. However, the EuroFIR prototype system should be valid for all type of nutrients, rather than only bioactive compounds.

It is clear that compilers should be able to precisely identify the nutrients to which the original data refer because misunderstanding the identity of the nutrient can dramatically modify the values to be assigned to a nutrient in a food.

Food description

- USDA and CSPO systems do not consider the criterion "food description".
- BASIS system is specific to plant foods, and is therefore precise for plant foods. State of maturity, season and place of production could also be added to describe plant foods more precisely.
- AFSSA system has rules common to all type of foods and also specific rules for manufactured, prepared foods and other specific rules for fresh "primary" foods. Manufactured prepacked foods and primary fresh foods have some common characteristics, but also some specificities: for example, asking for the region of production for breakfast cereals is probably not relevant, but asking the same question can be important for some fruits or fish. The. AFSSA system also includes the question "Is the common name of the food given?" although a definition is not given for "common name".

In USDA, AFSSA and CSPO systems, measurement of moisture is considered as part of sample handling. However moisture content can be used for food identification (to distinguish for example between coffee as a powder and as a beverage, when the state is not given), and it may also be appropriate to the "food description" category.

Sampling plan

- USDA and AFSSA systems grade sampling plan according to the method (probabilistic sampling or not) and the aim of the sampling (national representativeness or not). Nevertheless, we know from experience, that a real probabilistic sampling plan based on sound data is very difficult to prepare due to the lack of accessible data (e.g.: market shares). They also distinguish manufactured and prepacked food and primary foods, which are often produced and distributed across the country in different ways. For manufactured foods, production is often at national or international level, whereas for primary foods, production can depend on local supplies. Times and places of sampling are therefore often more important to consider for primary foods than for manufactured foods.

- CSPO system is based on the number of geographical areas studied, but season is not taken into account. In addition, CSPO also assigns an additional point when the sampling was made in Italy although, in some cases, foods that are sampled outside Italy can also be sold in Italy.

- BASIS assesses sampling based on the following question: "Is this a representative sample for this food plant?". It is important for a plant (or a piece of an animal) to sample the part that is consumed (e.g. for plants: root, leaf, stem, grain). This assessment is really more applicable to assessment of food description.

Number of samples

- AFSSA uses a more or less continuous scale (number of samples is multiplied by 2 to give a score) whereas CSPO uses a discontinuous scale (e.g.: more than 10 samples analysed = 3 points, 3 to 10 samples analysed = 2 points)

- BASIS does not assess the number of samples analysed.

Sample handling

- AFSSA rules are based on the USDA system: they concern storage, transportation, and homogenisation.

- USDA and CSPO systems consider "information on moisture" in sample handling.

- USDA, AFSSA and CSPO consider that the question concerning edible portion should be placed in this criterion. This issue is important, but there is discussion about the placement of this question because it may be better to include this assessment in the category "Food description". If nutrient values concern the non-edible part as well as the edible portion, the data cannot be used in a nutrient database, so it is better to assess this at the earliest stage possible (food description stage).

- BASIS makes this assessment based on 2 rules: "Post harvest/post purchase handling properly described and appropriate" (scoring a maximum of 10 points) and "Properly handled, frozen or dried immediately" (scoring a maximum of 10 points).

Analytical method

- AFSSA system defines "official methods" as ISO, AOAC or AOCS methods. However although some of these methods were previously the "best" for some nutrients in a matrix, after a period of time they can be outdated compared to more recent methods. An "official" method can also be applied in an inadequate matrix, leading to non-reliable results. Therefore, the use of an official list of approved methods for each component and matrix combination would be more appropriate for the EuroFIR prototype.

- CSPO system is specific to carotenoids and is based only on the application of HPLC. It can therefore not be extended to all nutrients in all matrices.

- BASIS gives a list of very interesting points to pay attention to when assessing analytical methodology. However this system is implemented by an 'expert' group and if applied more widely to food composition databases some likely problems would include:

1) Interpretation by compilers. Many compilers have no analytical background and would not know how to assess reproducibility, standard curves spiking, internal standard etc.

2) Lack of information on these aspects of analytical methodology.

Analytical performance / quality control

- AFSSA, BASIS and CSPO use different terms for this criterion:

- AFSSA: execution of the analytical method by the lab
- BASIS: analytical performance
- CSPO: quality control

Therefore common definitions are needed.

- AFSSA approach is mainly based on accreditation because:

- Many laboratories have obtained accreditation for the analysis of the nutrient composition of foods, and therefore the components of interest in a FCDB.

- Accreditation can be used to create a relatively “simple” rule (is the lab accredited for the analysis of this nutrient in the matrix concerned: yes or no?).

- Accreditation implies demonstration of analytical performance.

- Due to their insufficient analytical knowledge, compilers are in some cases not able to identify and interpret precise relevant information on analytical performance in laboratory reports ().

However, accreditation does not apply to data in scientific publications, which should provide sufficient information regarding analytical performance.

- CSPO has also found it very difficult to evaluate analytical performance and has therefore tried to simplify evaluation terminology, considering only accuracy and precision of the method. This system does not provide guidelines or rules for assessment.

- BASIS bases its assessment on reference materials and approved standards. It is necessary to define what “approved” standards are. How should compilers interpret the use of reference materials to evaluate analytical performance / quality control?

4. THE EUROFIR QUALITY ASSESSMENT SYSTEM

The EuroFIR system (*QI_GUIDELINES_Revised300608.doc*) has been developed taking into account comments derived by the detailed analyses of the existing systems. One of the aims is to allow data documented for the EuroFIR system to be comparable to data that evaluated using the USDA and BASIS systems. The development of the EuroFIR system has been conducted bearing in mind that quality evaluation within EuroFIR is currently carried out by a variety of compilers, with different levels of expertise in food chemistry. For this reason, rather than asking detailed information related to the analytical method, documents have been prepared by WP 1.3 (*see EuroFIR Technical Website*) for easy reference by compilers. These analytical guidelines describe, by food component, the appropriate analytical methods and steps to be applied to different food matrixes.

EVALUATION CATEGORIES

The following are therefore the resulting categories included in the EuroFIR system, with the relevant criteria to be evaluated (for a complete description of the system refer to the document *Oseredczuk et al. QI_GUIDELINES_Revised300608.doc)*. Answers to the criteria can be YES / NO / N/A (with the exception of the category "Number of analytical samples")

Food description

This is not a criterion in the USDA system, but it is a requirement for EuroFIR data interchange.

A. FOR ALL TYPES OF FOOD

Was the source of the food or of the main ingredient provided (best if scientific name included, cultivar/variety, genus/species, etc)?

Was the part of plant or part of animal clearly indicated?

If relevant, was the analysed portion described and is it clear if the food was analysed with or without the inedible part?

If relevant, was the extent of heat treatment provided?

If the food was cooked, were satisfactory cooking method details provided?
Was relevant information on treatment applied provided?
Was information on preservation method provided?
If relevant, was information on packing medium provided?
If relevant, was information about the origin of food provided?
If relevant, was the month or season of production indicated?
Was the moisture content of the sample measured and the result given?

B: FOR MANUFACTURED PREPACKED FOOD ONLY

Was the generic name provided (e.g. chocolate paste with hazelnuts)?
Was the commercial name provided (e.g. Nutella)?
Was brand provided (e.g. Ferrero)?
Was relevant information on consumer group/dietary use/label claim provided?

C: FOR HOME MADE DISHES OR FOODS SOLD IN RESTAURANTS

Was the complete name and description of the recipe provided?

Component identification

This is not a criterion in the USDA system, but it is a requirement for EuroFIR data interchange.

Is the component described unambiguously?
Is the unit unequivocal?
Is the matrix unit unequivocal?

Sampling plan

Was the sampling plan developed to represent consumption in the country where the study was conducted?
Was the number of primary samples >9?
If relevant, were samples taken during more than one season of the year?
If relevant, were samples taken from more than one geographical location?
If relevant, were samples taken from the most important sales outlet (supermarket, local grocery, street market, restaurant, household…)?
If relevant, was more than one brand (for manufactured pre-packed product) or more than one cultivar (for plant foods) or subspecies (for animal foods) sampled?

Number of analytical samples

Is the number of analytical samples 1, 2, 3, 4, or ≥ 5?

Sample handling

If relevant, were appropriate stabilization treatments applied (e.g. protection from heat/air/light/microbial activity), according to EuroFIR guidelines for analytical methods?
Were the samples homogenized?

Analytical method

Does the analytical method used in the source match the list of appropriate analytical methods given in the EuroFIR guidelines for analytical methods?
Are the key method steps appropriate for the method described, considering the EuroFIR guidelines as a reference?

Analytical quality control

Were analytical portion replicates tested?
Was the laboratory accredited for this method or was the method validated by performance testing?
If available, was an appropriate reference material used?

SUMMARY OF SCORING CATEGORIES

FOOD DESCRIPTION
Scoring: number of criteria answered positively * 5 / total number of criteria judged relevant
Possible scores: 1, 2, 3, 4 or 5 (after rounding)

COMPONENT IDENTIFICATION
Scoring: arbitrary (no calculation can be done)
Possible scores: 5 or 1 only (1 if one or more criteria are not satisfied)

SAMPLING PLAN
Scoring: arbitrary (no calculation can be done), some criteria may have more weight than others
Possible scores: 1, 2, 3, 4 or 5

NUMBER OF ANALYTICAL SAMPLES
Scoring: unambiguous
Possible scores: 1, 2, 3, 4 or 5

SAMPLE HANDLING
Scoring: arbitrary (no calculation can be done)
Possible scores: 5 or 1 only (1 if one or more criteria are not satisfied)

ANALYTICAL METHOD
Scoring: arbitrary (no calculation can be done)
Possible scores: 1, 2, 3, 4 or 5

ANALYTICAL QUALITY CONTROL
Scoring: arbitrary (no calculation can be done)
Possible scores: 1, 2, 3, 4 or 5 (3 is a minimum when the lab is accredited)

5. CONCLUSIONS

The USDA system is designed for limited use, for specific components and within a single country, while the EuroFIR system aims to be more generic and is designed for application to a wide range of foods and components.
The systems have in common the five main evaluation categories that are fundamental for allowing comparison between the quality of the data included in the two database systems. Food description and component identification are not assessed within the USDA system but are important within EuroFIR because data will be interchanged between countries and data compilers. The five other categories, related to sampling and analysis, are common to both systems so the scores assigned can be interchanged and could be used as an indication of strengths and weaknesses of documented data. However it should be noted that criteria within the categories are different and so the scores provided by the two systems are only an indication of quality and are not directly comparable.

6. REFERENCES

Greenfield H and Southgate DAT. Food composition data, production, management and use, 2nd edition, FAO, Rome 2003.

Holden JM, Eldridge AL, Beecher GR, Buzzard M, Bhagwat S, Davis CS, Douglass LW, Gebhardt ES, Haytowitz D and Schakel S. Carotenoid Content of U.S. Foods: An Update of the Database.J Food Comp Analysis 1999; 19:169-96.

Holden JM, Bhagwat SA, Patterson KY. Development of a multi-nutrient data quality evaluation system. J Food Comp Analysis (2002); 15: 339-348.

Holden JA, Bhagwat SA, Haytowitz DB, Gebhardt SE, Dwyer JT, Peterson J, Beecher GR, Eldridge AL, Balentine D.Development of a database of critically evaluated flavonoids data: application of USDA's data quality evaluation system. J Food Comp Analysis 2005; 18:829–44.

Holden_DQuality06.pdf (slides from Food Compo course, Bratislava, 2006)

Mangels AR, Holden JM, Beecher GR, Forman MR, and Lanza E. Carotenoid content of fruits and vegetables: an evaluation of analytic data. *J. Am. Diet. Assoc. 1994;* 93:284-96.

Oseredczuk M, Salvini S, Roe M, Møller A, Ireland J, et al. QI_GUIDELINES_Revised300608.doc

Oseredczuk M, Salvini S. QI_Comparisons_and_prototype_160806.doc

USDA System: Evaluation_Riboflavin.xls

USDA System: Evaluation_Vit K.xls