

Original Article

Validation study of the USDA's Data Quality Evaluation System

Seema A. Bhagwat*, Kristine Y. Patterson, Joanne M. Holden

Nutrient Data Laboratory, Beltsville Human Nutrition Research Center, Building 005, Room 107, BARC West, 10300 Baltimore Avenue, Agricultural Research Service, USDA, Beltsville, MD 20705-2350, USA

ARTICLE INFO

Article history:

Received 10 January 2008

Received in revised form 21 May 2008

Accepted 24 June 2008

Keywords:

Food composition database

Data quality

Data assessment

Data standardization

Data harmonization

System validation

ABSTRACT

The Nutrient Data Laboratory (NDL) of USDA conducted a validation study of the USDA Data Quality Evaluation System (DQES). The system evaluates the quality of analytical data on nutrients by rating important documentation concerning the analytical method, analytical quality control, number of samples, sampling plan, and sample handling and creates a "Quality Index" (QI) and "confidence code" (CC) for each nutrient and food. Objectives of the study were: (1) to measure the variability of ratings assigned by different evaluators, (2) to assess the objectivity of the "critical" questions of the DQES categories and (3) to test the robustness of the rating scale. Out of 39 individuals who participated in the International Postgraduate Course for Food Composition and four nutritionists from NDL, 37 completed the evaluation of a research article containing analytical data on vitamin K in lettuce, 25 evaluated an article containing data on catechin in black grapes, and 16 evaluated an article containing data on riboflavin in portabella mushrooms using the DQES. The various rating scores assigned by the participants were analyzed to assess the success of above objectives. The maximum score for each category was 20. There was a greater variation among individuals in the ratings for "number of samples" and "analytical quality control" categories than the other three categories. Overall seventy five percent of the participants assigned a confidence code of "C" to the vitamin K in lettuce, while 70% and 100% assigned confidence codes of "B" and "C" to catechin in black grapes and riboflavin in mushrooms, respectively. The wide range of scores on individual categories by first time evaluators emphasizes the importance of training for evaluators who score analytical values for inclusion in a food composition database. Clear documentation by the authors and training for evaluators to understand basic concepts related to data quality evaluation will be helpful. Future work will assess the assignment of equal rating points for all the categories. Modifications in "sampling plan" category to accommodate country and population size are necessary. The USDA Data Quality Evaluation System represents one of the first efforts to standardize and harmonize the evaluation of analytical data quality across the international food composition network.

Published by Elsevier Inc.

1. Introduction

Analytical data in many food composition databases (FCDBs) are obtained from direct chemical analysis, from published literature or both. Data from literature are collected by searching various databases like AGRICOLA (Agricultural Online Access), Food Science and Technology Abstracts (FSTA) and others that retrieve citations from agriculture, biology, environmental and

chemical sciences using key words for the nutrients of interest. Thus, data in the FCDBs are gleaned from diverse sources and may be of uneven quality and lacking in detailed supporting documentation. It is, therefore, of paramount importance to evaluate the quality of analytical data for its reliability and fitness for the purpose as well as to provide an indicator of data quality to guide the scientific community in the use of those data in various data applications.

In 1983 USDA's data quality evaluation procedure began as a manual system for evaluating analytical data for iron in foods (Exler, 1983). It was based on the evaluation of only three categories of attributes, each giving a rating that could range from 0 to 3 points: (1) documentation provided for the analytical method, (2) sample handling and appropriateness of the analytical method (treated as one category), and (3) analytical quality control. This model was subsequently expanded to include five

Abbreviations: NDL, Nutrient Data Laboratory; USDA, United States Department of Agriculture; DQES, Data Quality Evaluation System; QI, Quality Index; CC, confidence code; FCDB, food composition database; SP, sampling plan; SH, sample handling; AM, analytical method; AQC, analytical quality control; NS, number of samples; CRM, certified reference material; SRM, standard reference material.

* Corresponding author. Tel.: +1 301 504 0696; fax: +1 301 504 0632.

E-mail address: seema.bhagwat@ars.usda.gov (S.A. Bhagwat).

separate categories: (1) appropriateness of analytical method, (2) sample handling, (3) sampling plan, (4) number of samples, and (5) analytical quality control. The range of ratings for each category remained 0–3 points as before. This expanded system has been used to evaluate data for selenium (Bigwood et al., 1987; Holden et al., 1987; Schubert et al., 1987), copper (Lurie et al., 1989) and carotenoids (Holden et al., 1999; Mangels et al., 1993) in foods.

There has been increasing interest at the international level in assessing the quality of data in food composition databases, and it is now included as part of the International Postgraduate Course on the Production, Management, and Use of Food Composition Data (FoodComp) organized and now conducted by the Food and Agriculture Organization (FAO). As part of the training course, lectures on the assessment of data quality are combined with exercises for participants. As a result, in several countries, the departments or institutes responsible for developing food composition databases and which have had personnel participate in the postgraduate courses have adapted USDA procedures to their respective situations or developed their own systems for evaluating data quality. Over time, FoodComp participants and lecturers as well as others in the international community have generally agreed that the five basic categories of data attributes were useful for evaluation purposes and the concept of assigning ratings and confidence codes has been a significant advancement.

In 2002 USDA expanded the rating scale to 0–20 points for each category of evaluation to provide a continuous rating scale which assessed essential and specific aspects of the documentation. Questions covering critical material necessary for rating each category were revised to improve objectivity (Holden et al., 2002). For each data source (i.e., published information or manufacturer's data on nutrient content of food items to be evaluated) values for a food are rated separately on the five categories for each nutrient or food component in that food item. Then the ratings are summed to obtain a Quality Index (QI), which is a total score for the nutrient or food component in that food item, meant to denote the 'quality' of the nutrient value.

At USDA, it has been the practice to aggregate acceptable values for the same food item to obtain a mean value for that nutrient or food component across all sources of data. To determine the overall quality rating for the mean value, QIs from the individual sources in the final dataset were summed in earlier versions of the system (Mangels et al., 1993).

Subsequently, USDA modified the system procedures to assign quality ratings to the aggregated mean value for a nutrient or food component in a food item. Ratings for analytical method, sample handling, and analytical quality control for each nutrient/food component of the contributing data sources are now averaged across data sources and the average values then are assigned to the mean nutrient value for these three categories of evaluation. For the two remaining categories, sampling plan and number of samples, new ratings are generated based on the overall representativeness of the sampling for the aggregate (e.g., samples from a variety of geographic areas, etc.) and the total number of samples in the aggregate. These two new ratings for the sampling plan and the number of samples categories and the three averaged scores for the other three categories are then summed to make up a new QI for the nutrient or food component in the food item. The new QI generated in this manner for the aggregated data is used as the basis for the confidence code, which is an alphabetical code, i.e., A, B, C, or D and is used as an indicator of overall data quality. As a result, the process now generates and retains scores for each category of evaluation within the dataset to provide more detailed information about the quality of the dataset.

The new Data Quality Evaluation System (DQES) has been used in the evaluation of data to be included in USDA's Special Interest

Databases such as Flavonoids (Holden et al., 2005), Department of Agriculture (2004), ORAC, 2007, and is currently being applied to the update of USDA's isoflavones database.

However, the variability in ratings assigned by different evaluators using similar algorithms has never been assessed. Thus the objectives of the present study are to: (1) assess the objectivity of the "critical" questions that form the basis of evaluating the DQES system categories, (2) measure the variability in ratings assigned by different evaluators, and (3) test the robustness of the rating scale. Results of this study will facilitate further modification of the system, if necessary, and implementation at several levels of database development.

2. Methods and materials

Post-Graduate FoodComp courses are conducted in different regions of the world for database compilers and analysts as well as data users. These courses are currently coordinated by Food and Agriculture Organization (FAO), the United Nations University (UNU), European (2005), other regional and national governments, non-government organizations, and other private sector groups. More than 350 participants from approximately 50 countries have been trained since 1992. The syllabus of the course includes lectures and question/answer sessions on data quality evaluation. Participants who attended recent courses conducted at Wageningen, The Netherlands, in 2005, Bratislava, Slovakia, in 2006, and Hyderabad, India, in 2006, as well as three nutritionists at the Nutrient Data Laboratory (NDL)/USDA, reviewed two to three published articles from scientific journals to evaluate aspects of the DQES. Thirty-four participants and three nutritionists at the NDL evaluated the data quality on leaf lettuce from one article on phylloquinone (vitamin K) content in vegetables, fruits, and berries for this exercise (Koivu et al., 1997). Twenty-three participants who attended the courses in Bratislava and India and two nutritionists at the NDL evaluated black grapes from a second article on catechin content of Bulgarian fruits (Tsanova-Savova et al., 2005). A third article on the riboflavin content in portabella mushrooms (Esteve et al., 2001) was evaluated by thirteen participants who attended the course at Wageningen in 2005 and three nutritionists at the NDL. These three articles included a variety of elements of the five categories used for the data quality evaluation. Sampling plans in the articles ranged from very simple (from riboflavin article) to complex plan based on agricultural statistics (for catechin article). All the articles documented analytical methods in details, while the determination of the number of samples analyzed required careful interpretation. Therefore, it seemed that these articles could offer a good exercise in understanding the basic concepts involved in the data quality evaluation.

Electronic templates of the evaluation system were provided to the participants. These templates included lists of critical questions for each of the five evaluation categories: sampling plan (SP), sample handling (SH), analytical method (AM), analytical quality control (AQC), and number of samples (NS). The evaluators were required to read the articles carefully, noting details of the categories which have been published by the authors. Then, each evaluator responded to the specific questions by selecting from the possible answers presented in successive drop-down menus. The ratings and relative confidence in the published estimates are largely dependent upon the adequacy of published documentation. Also, the evaluator's knowledge of food composition processes is essential to the assessment of data quality. Information about the facets of the evaluation categories were described earlier by Holden et al. (2002). However, the procedures described below emphasize the main areas where points are assigned in the system.

IF	THEN
Regions (1– 4) sampled in Country X?	Range of possible points (1 – 10)
Cities (1-≥3) sampled within a region?	Range of additional points (1-3)
Locations (1-≥2) sampled per city?	Range of additional points (0-1)
Lots (1-≥2) sampled per location?	Range of additional points (0-1)
Seasons (1-≥2) sampled?	Range of additional points (0-1)

Fig. 1. Example of DQES evaluation of non-probability sampling plan showing how points are assigned (maximum of 16) for foods sampled in country X.

The SP category rates the information regarding the representativeness of the food samples analyzed. Fig. 1 illustrates the information sought and the allocation of points to the questions for this category. To achieve national representativeness of the nutrient value(s) it is desirable to sample widely across several regions of a country. Therefore, within reasonable limits, the higher the number of regions sampled, the more representative the sample (no. of units) will be.

The system category for assessing AM emphasizes that careful planning by the analytical team is necessary before actual analyses can be started. The published plan should involve selecting appropriate analytical methods and establishing an analytical quality control program (Fig. 2). AMs are nutrient-specific and are evaluated in two parts. In part one, the selected analytical method is evaluated according to the question list previously established by the experts in that area and qualifies for a maximum of eight points. The analyst(s) should demonstrate the accuracy and precision of the AM as executed in their own laboratory. Fig. 3 demonstrates the evaluation of the second part, the rating of the execution of the method (validation), which is awarded a maximum of 12 points. The analytical method should be validated before starting the chemical analysis of experimental samples using certified/standard reference material (CRM/SRM), comparing results with another laboratory or another analytical method.

The description of the analysis of a reference in-house quality control material QCM (in-house) (Emons, 2006), should be provided. This QCM (in-house) material should be analyzed along with the analytical samples throughout the duration of the analysis described in the source. It is used to monitor precision during the on-going execution of the study and sample analysis. Fig. 4 illustrates the assignment of points for the AQC category.

Determining the number of samples analyzed for the NS category can be complex or confusing since it requires ascertaining the number of individual sample units analyzed. The majority of publications do not report clearly this information. Fig. 5 explains how to determine the number of individual samples analyzed for this category.

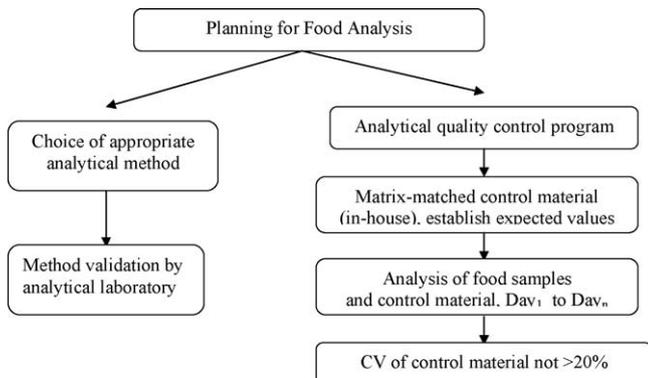


Fig. 2. Planning schema for analytical method and quality control in food analyses.

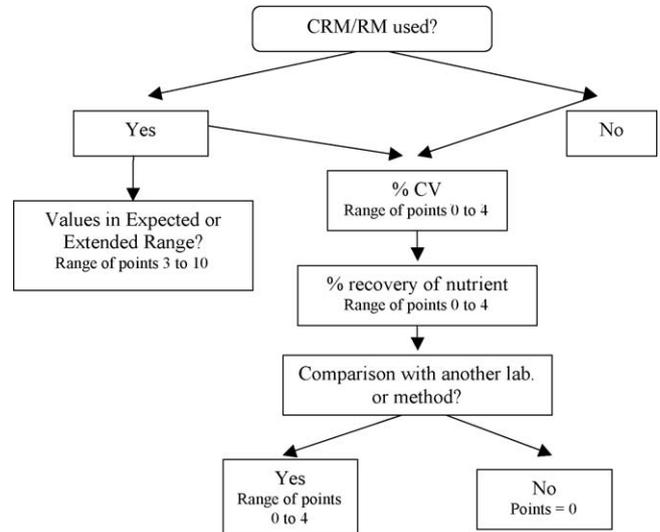


Fig. 3. Example of DQES evaluation of the execution of the analytical methodology by the laboratory showing how points are assigned (maximum of 12).

In addition to rating the five aforementioned categories, the template also includes a section to collect information about the general description of the food analyzed. However, in the current version of the USDA system this section was not rated. This additional category of evaluation was suggested by Møller et al. (2007) to stress the importance of detailed and complete food description. Critical questions included in the sections on SP, SH, AQC, and NS are usually the same for all nutrients included in a data source, but the AM section for each source relates exclusively to the nutrients analyzed; in this exercise, specifically vitamin K, catechin, and riboflavin.

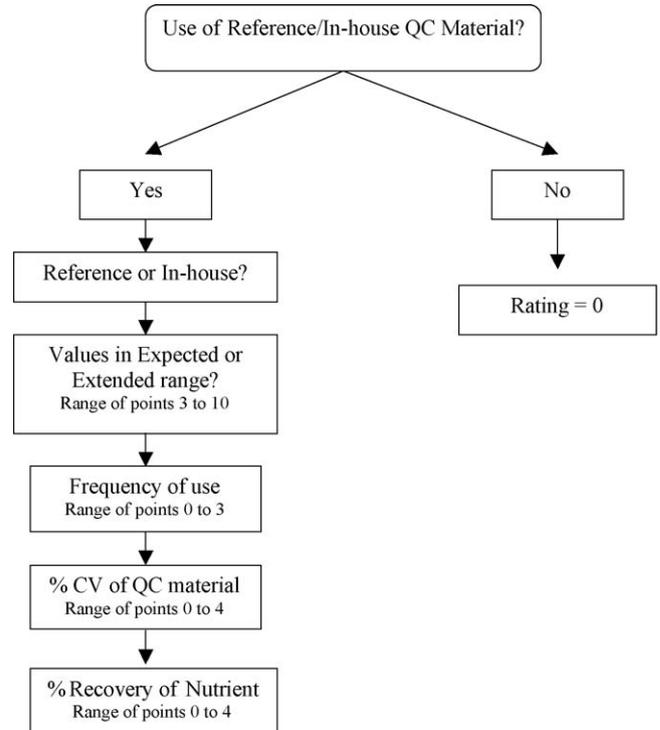


Fig. 4. Example of DQES evaluation of the analytical quality control by a laboratory showing how points are assigned (maximum of 20).

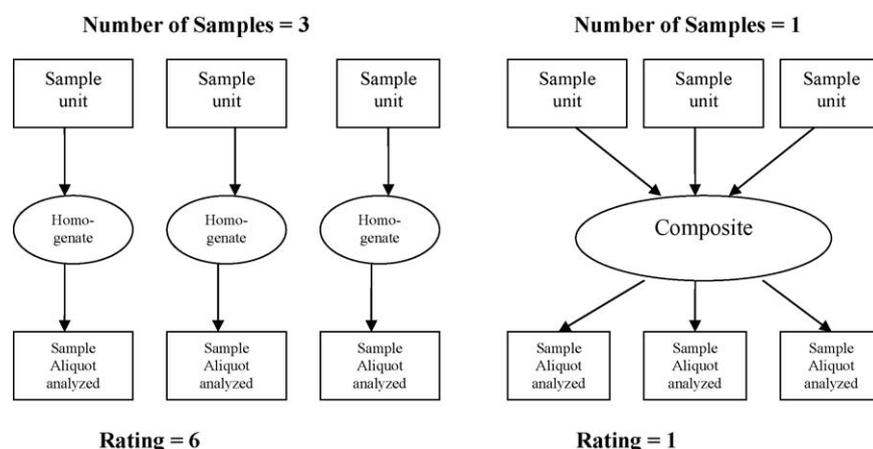


Fig. 5. Example of DQES determination of the number of samples.

For this exercise, multiple choices were provided to facilitate answering each question in all the five categories. The DQES system assigns points to each response, and then sums the points to get the rating for each category. The point range for each category is 0–20. Occasionally for an evaluation category for a specific publication the sum of rating points may exceed 20 points. If this occurs the recorded rating will default to a maximum of 20 points. QIs were calculated by summing the ratings of all the five categories for this exercise. A maximum QI score is 100. Based on the resulting QI scores, CCs were assigned for each source. Table 1 provides ranges for assignment and the meaning of CCs.

The completed electronic evaluation forms were submitted by the evaluators and analyzed to assess the responses for each article. Evaluation category ratings for selected food from each article were tabulated in Excel files. Minor preliminary adjustments were made to correct for inconsistent responses. For example, ratings for AQC were corrected to “0” rating if an evaluator provided conflicting answers in a series of questions; e.g., an evaluator indicated that no QCM (in-house) material was used, and then answered the next questions instead of skipping the rest of the questions as directed (Fig. 4). Evaluation responses for the vitamin K and catechin articles by one participant seemed unreasonable and were excluded from the dataset. Next, mean and median ratings for each category, as well as ranges of QIs and CCs were calculated.

Table 1
Assignment and meaning of confidence codes.

Quality Index value range	Confidence code	Meaning of confidence code
100–75	A	The user can have considerable confidence in this value
74–50	B	The user can have confidence in this value; however some problems exist regarding the data on which the value is based
49–25	C	The user can have less confidence in this value due to limited quantity and/or quality of data
<25	D	There are significant problems with the value related to limited quantity and/or quality of data

Table 2
Mean, median and ranges of ratings assigned to five categories of evaluation by participants.

Nutrient	SP ratings mean/ median (range)	SH ratings mean/ median (range)	AM ratings mean/ median (range)	AQC ratings mean/ median (range)	NS ratings mean/ median (range)
Vitamin K	4.8/5 (2–6)	11.3/12 (6–17)	15/15.5 (7–18)	6/5 (0–17)	6.9/3 (1–20)
Catechin	13.1/13.5 (5–17)	14.9/15 (12–20)	10.2/10 (7–15)	2.4/0 (0–17)	14/14 (14)
Riboflavin	2.3/2 (2–6)	9.3/10 (7–17)	14.2/16 (5–17)	2.5/0 (0–19)	11.3/12 (1–12)

3. Results and discussion

Responses of evaluators to each question in each category and resulting ratings were studied carefully to assess the degree of consistency in responses among reviewers. Results for each category are reported in Table 2. The maximum possible rating for each evaluation category was 20 points.

3.1. Sampling plan

The mean values for SP ratings for vitamin K, catechin and riboflavin were 4.8, 13.1, and 2.3, respectively (Table 2, Fig. 6).

The question on the method of sampling, i.e., probability vs. non-probability, appeared to not be well understood by the evaluators. This was identified by the response to the question, “Was the sampling plan based on a statistical probability model?” The response given by more than half the participants (34 out of 78) was “unknown”.

Nationwide, multi-region sampling schemes are given higher ratings by the USDA than other options. Current sampling plan criteria are suitable for large countries like the United States or Brazil, but for smaller countries samples acquired from only a few locations could be nationally representative. Therefore, alterations to the DQES may be necessary for smaller countries taking into consideration percent consumption relative to population centers, total size of population, and country size (land mass vs. population).

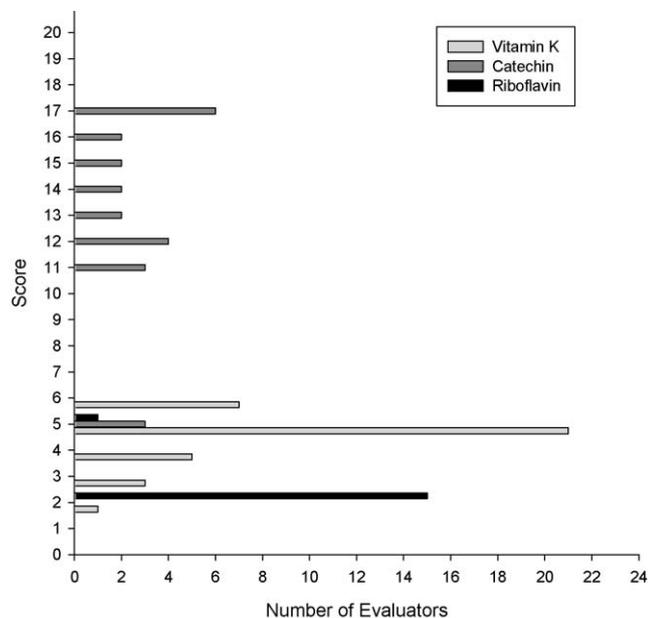


Fig. 6. Sampling plan scoring for articles on vitamin K, catechin, and riboflavin by various DQES participating evaluators.

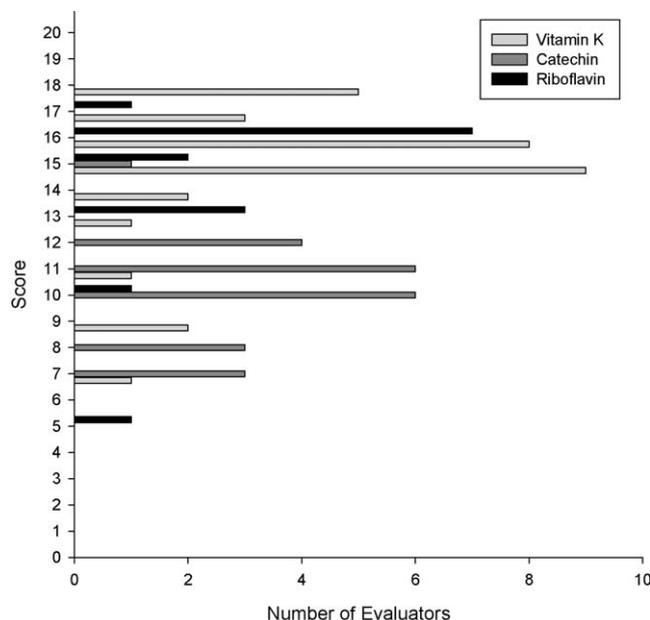


Fig. 8. Analytical method scoring for articles on vitamin K, catechin, and riboflavin by various DQES participating evaluators.

Points assigned to each question would also need reconsideration/adjustment to reduce the emphasis on higher number of regions covered by sampling, particularly in smaller countries.

3.2. Sample handling

The mean values for SH ratings for vitamin K, catechin, and riboflavin were 11.3, 14.9, and 9.3, respectively (Table 2, Fig. 7).

The question concerning the verification of homogenization procedures for analytical samples produced ambiguous answers from participants. Some participants seemed to have a lack of knowledge about the process for verifying homogeneity of sample material. If homogeneity is not verified then reported values may

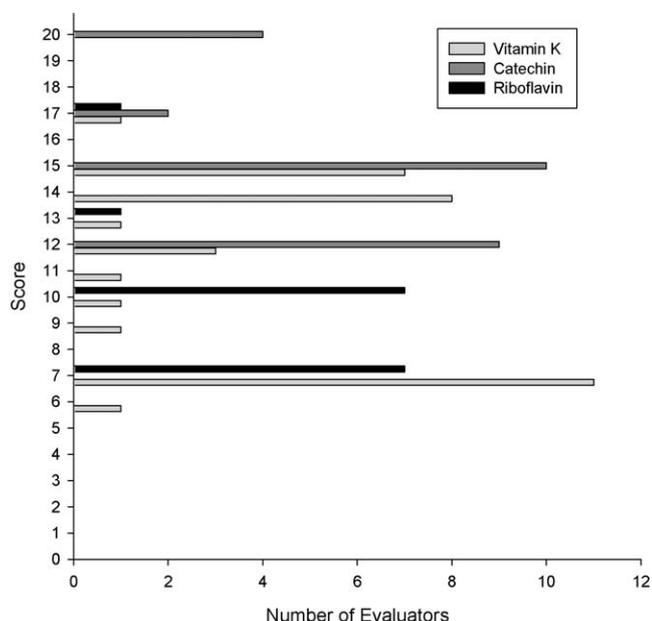


Fig. 7. Sample handling scoring for articles on vitamin K, catechin, and riboflavin by various DQES participating evaluators.

not be reliable. Either the report of a standard deviation for the replicate analyses of the same homogenate or a more general statement about the development of standard protocols could meet the requirement for obtaining the points allotted to this question. Standard protocols are required for every step of the sample handling process, including storage temperature, description of edible portion (e.g., with/without skin), and homogenization (e.g., equipment, storage of homogenates, freeze-drying). To receive the total rating points the source article should provide the details or mention a previous reference which contains the information.

3.3. Analytical method

The mean values for AM ratings for vitamin K, catechin, and riboflavin were 15.0, 10.2, and 14.2, respectively (Table 2, Fig. 8). Some experience and/or training in chemistry are essential to answer questions in this category. The provision of definitions for technical terms like CRM/SRM and values associated with them may assist evaluators.

3.4. Analytical quality control

Questions in the AQC category evaluate accuracy and precision in the day-to-day execution of an analytical method. This category was the most difficult to evaluate in the exercise and requires understanding of the underlying concepts and reasoning for its necessity. The mean AQC for vitamin K, catechin, and riboflavin were 6.0, 2.4, and 2.5, respectively (Table 2, Fig. 9).

This category is the one most neglected by analysts; little or no information is usually reported in publications. Certified or Standard Reference Materials (CRM/SRM), if available, are used to validate the analytical method. However, those materials tend to be very expensive for use as quality control (QC) materials. For any analyses, particularly those performed over a period of time, QCM (in-house) material is required to verify consistent performance of the analytical method, including instrument performance. An in-house material similar in matrix to the analytical samples should be prepared and used with every batch of the samples analyzed.

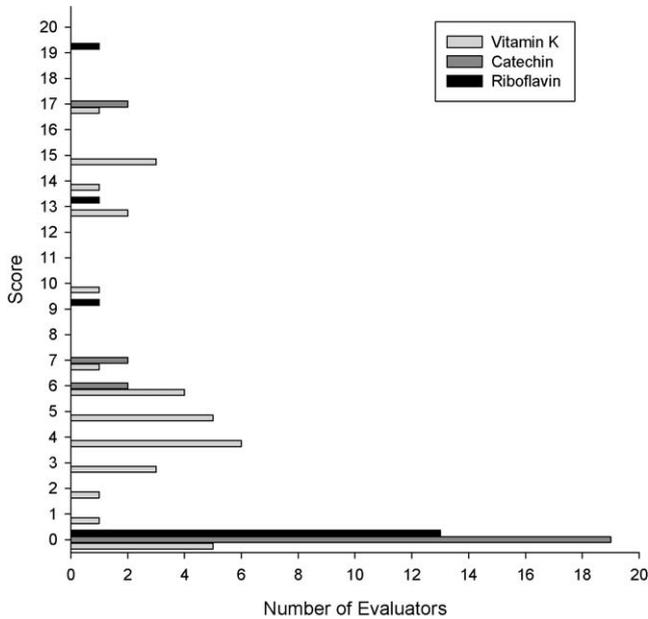


Fig. 9. Analytical quality control scoring for articles on vitamin K, catechin, and riboflavin by various DQES participating evaluators.

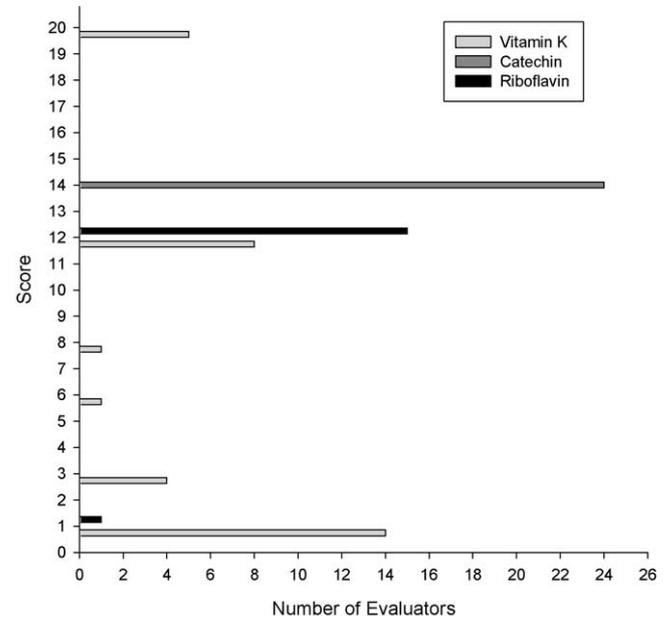


Fig. 10. Number of samples scoring for articles on vitamin K, catechin, and riboflavin by various DQES participating evaluators.

Deviations from the reference values established a priori for this material are used to assess potential problems in the execution of the method and %relative standard deviations (%R.S.D.) for the replicate analytical values of this material over a period of time verify the precision of the method. The %R.S.D. includes variability attributable to analytical instrument, the technician performance, and to the method itself.

3.5. Number of samples

The mean values for NS category ratings for vitamin K and riboflavin were 6.9 and 11.3, respectively. The range of scores, 1–20, for vitamin K indicates that it was difficult for evaluators to determine how many samples were analyzed. However, a rating of 14 was unanimous for catechin because of clear reporting by the authors (Table 2, Fig. 10).

This category was one of the two most difficult categories to evaluate, the other being the AQC category. Fig. 5 illustrates the concept of number of individual samples analyzed. If samples procured from various locations are mixed to make a single composite, then the number of samples analyzed remains 1 and achieves the rating of 1 only. This will be true despite the fact that more than one sub-sample (sample aliquots) from the single composite were analyzed independently. Under these circumstances SP category will get a higher rating, but not the NS category. This category rates the number of “analytical” samples that were analyzed; the same rating would be assigned to a sample prepared from a homogenate of one individual sample unit or be a composite of several units.

3.6. Quality indices and confidence codes

Table 3 describes the ranges of QIs and resulting CCs for the three nutrient values for the selected foods evaluated in the respective publications, i.e. data sources. Twenty-five percent of the 36 evaluators assigned a CC of B (QI range 51–70), while 75% assigned a CC of C (QI range 26–48) for the vitamin K value for leaf lettuce. Seventy percent of 24 participants who evaluated the catechin value assigned a CC of B (QI range 51–70) and 30%

assigned a CC of C (QI range 46–48) for black grapes, while 100% assigned a CC of C (QI range 31–47) to the riboflavin value of portabella mushrooms. The range of QIs for a CC of C for catechin was very narrow (46–48) and very close to the lower value of the range for CC of B, which is 50 (see Table 1). Although the QI range for a CC of B for catechin was 51–70, only 3 values out of 16 were above 61 while 13 were between 51 and 58. This observation demonstrates that the differences in the ratings of catechin values for black grapes were not striking and most of the ratings were closer to a CC of B. All the participants assigned a CC of C for riboflavin values for portabella mushrooms. The slightly larger difference in evaluations for vitamin K values in leaf lettuce can be attributed to the ratings for AQC category which, as mentioned above, was the most difficult to evaluate.

3.7. General issues

In addition to the observations regarding the need for modification of the system itself already mentioned, one of the general issues highlighted by this exercise was related to the distribution of points. The present system regards all the five categories as equally important and gives a maximum of 20 points to each category. The AM category should perhaps get a larger share of the total 100 points, followed by SP, SH, AQC, and NS. In addition, the use of electronic templates will require selected modification/expansion to sampling modules. Basic software enhancements could help eliminate ambiguous responses from

Table 3
Quality indices and confidence codes^a.

Nutrient	Quality Index (Ranges)	Confidence codes (% participants)
Vitamin K	51–70	B (25%)
	26–48	C (75%)
Catechin	51–70	B (70%)
	46–48	C (30%)
Riboflavin	31–47	C (100%)

^a Maximum possible Quality Index = 100.

participants. For example, if the answer to the first question in the AQC section, whether the QCM (in-house) material was analyzed or not, is “No”, the category should get an automatic “0” rating. Some participants suggested making the % CV and %R.S.D. ranges more specific intervals, e.g., 10–14%, 15–19% and $\geq 20\%$, instead of $\leq 10\%$, $\leq 15\%$ or $\leq 20\%$. They reported that the “less than ranges” were confusing. These changes have already been incorporated in the system.

It is important that the evaluators answer all the necessary questions in all the categories. Eight out of 37 participants who evaluated vitamin K, and one each out of 25 and 16 participants who evaluated catechin and riboflavin, respectively, did not complete all questions in the five categories. Therefore, those ratings could not be included in the final compilation and assignment of CCs. It is also important that the authors/analysts provide clearly written detailed documentation of food sample description (e.g. fruits with/without skin, cooking method for cooked foods, all or drained solid only for canned foods, etc.), sample handling from the procurement to sample preparation for analysis, analytical method details (% recoveries, %CV, etc.), analytical QC details, and number of individual samples analyzed. Frequently, it is not clear if the mention of replicates implies repeated analyses of the same sample homogenate or homogenates of different distinct individual analytical samples.

3.8. Future recommendations

As a result of these exercises, several suggestions for future consideration were identified: (1) create a single web site for multiple users. Multiple evaluators can use standardized software to evaluate articles containing composition data for foods. (2) Develop an international reservoir of evaluated articles. Scientists from various research centers could access articles of common interest which have already been evaluated. There would be less need for multiple scientists to evaluate the same articles. (3) Retain individual scores for each category. Individual scores for each category can indicate the strengths and weaknesses of each article. The user of the data would then be able to distinguish between the qualities of two or more articles which have received similar QIs. (4) Publish critical questions for each class of nutrients or other compounds. If critical questions are published then the analysts/authors and data compilers will be more aware of data quality criteria. (5) Create a glossary of terms. A glossary of relative terms is needed to clarify any misunderstanding on the part of the compiler. (6) Maintain a complete list of reference materials and values. Evaluators can use a comprehensive list of quality control materials to get further information to be used in the validation of the reported method. (7) Expand sampling plan options. This will permit a higher rating for sampling in smaller countries where extensive local sampling may have been performed. (8) Develop a guide for using data quality evaluation scores and indicators (CCs). This guide can assist data users to select specific composition values fit for their purpose.

4. Conclusions

Participants with various backgrounds and from more than 35 countries participating in post-graduate FoodComp courses conducted in 2005 and 2006 completed exercises to determine if the new DQES system could provide meaningful information on data quality in an international setting. It can be concluded from this

validation study that generally consistent results can be obtained by many evaluators for data quality evaluation with a few modifications in the templates and clarifications of technical terms in the questions. Clear and detailed documentation in publications will facilitate the evaluation process by removing ambiguities.

This was the first attempt to evaluate the sources of data in publications by most of the participants. Results of this study have demonstrated that more training for compilers, analysts, and users of food composition data is needed to increase understanding of basic concepts related to the evaluation of data quality. As demonstrated in this exercise, some compilers had more difficulty than others and thus, would require more training in these topics. Compilers may require training or/and experience in analytical techniques and the generation of data in order to evaluate data quality. Modifications for the SP category have been proposed to address diverse populations and localities and the size and distribution of population, particularly in the countries of smaller size.

References

- Bigwood, D.W., Heller, S.R., Wolf, W.R., Schubert, A., Holden, J.M., 1987. SELEX: An expert system for evaluating published data on selenium in foods. *Analytica Chimica Acta* 200, 411–419.
- Emons, H., 2006. The ‘RM family’—Identification of all its members. *Accreditation and Quality Assurance* 10, 690–691.
- Esteve, M.-J., Farré, R., Frigola, A., Garcia-Cantabella, J.-M., 2001. Simultaneous determination of thiamin and riboflavin in mushrooms by liquid chromatography. *Journal of Agricultural And Food Chemistry* 49, 1450–1454.
- European Food Information Resource Network. (2005/2009). <http://www.eurofir.net>.
- Exler, J. (1983). Consumer Nutrition Division, Human Nutrition Information Service, USDA, [Home Economics Research Report Number 45].
- Holden, J.M., Bhagwat, S.A., Haytowitz, D.B., Gebhardt, S.E., Dwyer, J.T., Peterson, J., et al., 2005. Development of a database of critically evaluated flavonoids data: Application of USDA’s data quality evaluation system. *Journal of Food Composition and Analysis* 18, 829–844.
- Holden, J.M., Bhagwat, S.A., Patterson, K.Y., 2002. Development of a multi-nutrient data quality evaluation system. *Journal of Food Composition and Analysis* 15, 339–348.
- Holden, J.M., Eldridge, A.L., Beecher, G.R., Buzzard, I.M., Bhagwat, S., Davis, C.S., et al., 1999. Carotenoid content of U.S. foods: An update of the database. *Journal of Food Composition and Analysis* 12, 169–196.
- Holden, J.M., Schubert, A., Wolf, W.R., Beecher, G.R., 1987. A system for evaluating the quality of published nutrient data: Selenium, a test case. *Food and Nutrition Bulletin* 9 (Suppl.), 177–193.
- Koivu, T.J., Piironen, V.I., Henttonen, S.K., Mattila, P.H., 1997. Determination of phyloquinone in vegetables, fruits, and berries by high-performance liquid chromatography with electrochemical detection. *Journal of Agricultural and Food Chemistry* 45, 4644–4649.
- Lurie, D.G., Holden, J.M., Schubert, A., Wolf, W.R., Miller-Illi, N.J., 1989. The copper content of foods based on a critical evaluation of published analytical data. *Journal of Food Composition and Analysis* 2, 298–316.
- Mangels, A.R., Holden, J.M., Beecher, G.R., Forman, M.R., Lanza, E., 1993. Carotenoid content of fruits and vegetables: An evaluation of analytic data. *Journal of the American Dietetic Association* 93, 284–296.
- Møller, A., Unwin, I.D., Becker, W., Ireland, J., 2007. EuroFIR’s food databank system for nutrients and bioactives. *Trends in Food Science and Technology* 18, 428–433.
- Schubert, A., Holden, J.M., Wolf, W.R., 1987. Selenium content of a core group of foods based on a critical evaluation of published analytical data. *Journal of the American Dietetic Association* 87, 296–299.
- Tsanova-Savova, S., Ribarova, F., Gerova, M., 2005. (+)-Catechin and (–)-epicatechin in Bulgarian fruits. *Journal of Food Composition and Analysis* 18, 691–698.
- U.S. Department of Agriculture, Agricultural Research Service. (2004). *USDA database for the Proanthocyanidin content of selected foods*. Nutrient Data Laboratory home page: <http://ars.usda.gov/nutrientdata>.
- U.S. Department of Agriculture, Agricultural Research Service. (2007). *USDA database for the oxygen radical absorbance capacity (ORAC) of selected foods*. Nutrient Data Laboratory home page: <http://ars.usda.gov/nutrientdata>.