Journal of Food Composition and Analysis 24 (2011) 629–636

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

journal homepage: www.elsevier.com/locate/jfca

Original Article

Brazilian flavonoid database: Application of quality evaluation system

Elizabete Wenzel de Menezes [1,*], Nelaine Cardoso Santos, Eliana Bistriche Giuntini, Milana C.T. Dan, Maria Ines Genovese, Franco M. Lajolo [1]

Depto. Alimentos e Nutrição Experimental, Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, 05508-000 São Paulo, SP, Brazil

ARTICLE INFO

Article history:
Received 9 December 2009
Received in revised form 5 July 2010
Accepted 7 September 2010
Available online 16 November 2010

Keywords:
Flavonoids
Data quality evaluation system
Food composition
Brazilian flavonoid database
Brazilian Food Composition Database
TBCA-USP
Biodiversity

ABSTRACT

Much information on flavonoid content of Brazilian foods has already been obtained; however, this information is spread in scientific publications and non-published data. The objectives of this work were to compile and evaluate the quality of national flavonoid data according to the United States Department of Agriculture's Data Quality Evaluation System (USDA-DQES) with few modifications, for future dissemination in the TBCA-USP (Brazilian Food Composition Database). For the compilation, the most abundant compounds in the flavonoid subclasses were considered (flavonols, flavones, isoflavones, flavanones, flavan-3-ols, and anthocyanidins) and the analysis of the compounds by HPLC was adopted as criteria for data inclusion. The evaluation system considers five categories, and the maximum score assigned to each category is 20. For each data, a confidence code (CC) was attributed (A, B, C and D), indicating the quality and reliability of the information. Flavonoid data (773) present in 197 Brazilian foods were evaluated. The CC "C" (as average) was attributed to 99% of the data and "B" (above average) to 1%. The main categories assigned low average scores were: number of samples; sampling plan and analytical quality control (average scores 2, 5 and 4, respectively). The analytical method category received an average score of 9. The category assigned the highest score was the sample handling (20 average). These results show that researchers need to be conscious about the importance of the number and plan of evaluated samples and the complete description and documentation of all the processes of methodology execution and analytical quality control.

1. Introduction

The scientific community has significant interest in the quantification and characterization of different flavonoids present in Brazilian foods; however, the information is dispersed in publications or in laboratory internal data and thesis (Genovese and Lajolo, 2002; Arabbi et al., 2004; Santos, 2005; Matsubara and Rodriguez-Amaya, 2006a,b; Genovese et al., 2007; Rosso et al., 2008; Santos, 2009). There are two main reasons for generating this information. The first one is due to the anti-inflammatory, antioxidative and antimicrobial properties of these bioactive compounds and their possible effects on decreasing the risk for non-transmissible chronic diseases (NTCD) (Kris-Etherton et al., 2004; Gry et al., 2007; Denny and Butriss, 2007). The second one is related to the Brazilian biodiversity of plant foods, which involves the necessity of knowing the content and type of flavonoids not only in conventional foods but also in region-specific ones (Toledo and Burlingame, 2006; Menezes, 2009; Burlingame et al., 2009).

Researchers from other regions like North America and Europe developed databases of bioactive compounds or specific compounds aiming to unite data that would allow a real evaluation of ingestion of these substances by the population. The Nutrient Data Laboratory (NDL) of the USDA made isoflavones data available in 1999 (Release 1) and in 2008 (Release 2) (USDA, 2008). The USDA Special Interest Database for flavonoid content of selected foods was introduced in the NDL website in March, 2003. This database contained values of 225 foods from different countries and all flavonoid data were critically evaluated according to the USDA's data quality evaluation system (USDA DQES) (Holden et al., 2005). After observing great variability between values of flavonoid data from Europe and other countries in relation to those from the US, researchers from the NDL warned about the necessity of analyzing foods that are commercialized and consumed in the country,

Abbreviations: QI, Quality Index; CC, confidence code; NDL, Nutrient Data Laboratory; BRASILFOODS, Brazilian Network of Food Data Systems; TBCA-USP, Brazilian Food Composition Database-USP; LATINFOODS, Latin American Network of Food Data Systems; INFOODS, International Network of Food Data Systems; EuroFIR, European Food Information Resource project; CV, coefficient of variation; RM, Reference Material; CRM, certified reference material; SRM, standard reference material; NTCD, non-transmissible chronic diseases; USDA-DQES, United States Department of Agriculture's Data Quality Evaluation System.

* Corresponding author at: Universidade de São Paulo (USP), FCF, Depto. Alimentos e Nutrição Experimental, Bl. 14, Av. Prof. Lineu Prestes 580, 05508-000 São Paulo, SP, Brazil. Tel.: +55 11 3091 36 24; fax: +55 11 3815 44 10.
E-mail address: wenzelde@usp.br (E.W. de Menezes).
[1] Coordinators of TBCA-USP and BRASILFOODS.

0889-1575/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jfca.2010.09.004

which resulted in the increase in available data (Release 2.1) (USDA, 2007).

The EuroFIR (European Food Information Resource) project started the development of harmonized and standardized food composition databases in Europe. The EuroFIR-BASIS database combines critically assessed compositional and biological effects data, including all the most important bioactive groups of plant origin (Gry et al., 2007).

Data from two databases (EuroFIR-BASIS database and USDA Special Interest Database for flavonoids) had their quality evaluated; however, once the systems differ in relation to the evaluation criteria adopted, one same component from an article can be differently classified depending on the system used. The EuroFIR-BASIS critical evaluating scoring system was based in six key components (plant/food description, sampling plan, sample handling, component description, analytical method and analytical performance) (Gry et al., 2007). The evaluation of each component consists of selecting a Yes or No response or assigning a score from one to five. The strength and weakness of the study are expressed through the response to each component and an overall numerical score is automatically calculated from the responses, presenting the user with an immediate overview of the data quality. Other tools have been created by the EuroFIR project to assure the quality of the data compilation process (Westenbrink et al., 2009; Castanheira et al., 2007).

The USDA DQES for flavonoids (Holden et al., 2005) was developed using procedures of multi-nutrient data evaluation module (Holden et al., 2002) and several previous experiences of data quality evaluation (Holden et al., 1987; Mangels et al., 1993). All data of each article were evaluated for five quality categories (number of samples, sampling plan, sample handling, analytical method and analytical quality control). Within each category, specific questions were defined to describe the critical steps necessary for achieving accurate and representative data. The scores for the five categories for each compound were summed to yield a Quality Index (QI). According to the data QI, the confidence code (CC) is assigned, indicating the relative quality of the data and the reliability of the mean. The USDA DQES was validated by Bhagwat et al. (2009). The evaluation of data quality helps to set priorities and further identifies the foods to be analyzed as well as orientate analysts to generate high quality data on flavonoids in foods.

The objectives of this work were to compile and evaluate the quality of national flavonoid data according to the USDA DQES for future dissemination of this information in the TBCA-USP.

2. Methodology

2.1. Description of data compilation and database

Data from foods that are produced and commercialized in Brazil, present in scientific publications and non-published data (thesis) were compiled. In order to facilitate data compilation and to guarantee data harmonization, the Form for Compilation of Food Composition Data, developed by BRASILFOODS (Menezes et al., 2005), was updated. The spreadsheets for identification of foods and analytical quality control were not modified. The INFOODS guidelines (Truswell et al., 1991) to describe foods were adopted with modifications made by LATINFOODS (FAO, 1995; FAO/LATINFOODS, 2004; Menezes et al., 2005). However, a spreadsheet for flavonoid data was created according to the flavonoid subclasses and the most abundant compounds (28) in foods, including: flavonols – isorhamnetin, kaempferol, myricetin, quercetin; flavones – apigenin, luteolin; isoflavones – genistein, daidzein, glycitein; flavanones – eriodictyol, hesperetin, naringenin; flavan-3-ols – catechin and gallic acid esters of catechin, epicatechin and gallic acid esters of epicatechin, theaflavin and gallic acid esters of theaflavin, thearubigin; anthocyanidins – cyanidin, delphinidin, malvidin, pelargonidin, peonidin, petunidin. The INFOODS tagnames for flavonoids were used to improve data interchange (INFOODS/FAO, 2009). The spreadsheet completed for flavonoid data represents the profile of information as it is presented in the Brazilian flavonoid database.

The Brazilian flavonoid database presents the content of each flavonoid compound as mg/100 g of fresh weight of edible portion (expressed as aglycons) with the respective standard deviation or variation and the content of moisture as g/100 g of edible portion of food. When data on moisture was not provided in the article, the author was contacted or the information of a similar food was taken from the TBCA-USP or a new sample of the same food was analyzed. Data provided as dry weight were transformed into fresh weight, resulting in loss of information regarding the standard deviation or variation. Mostly the authors had to be contacted to provide additional information, such as data in the form of graphics, values expressed as %, total value of a subclass component without its separation, among others. Values for beverages, foods for special diets, infant formulas and others were adjusted by their respective specific gravities and were reported as mg/100 g. In the case of teas, flavonoids were presented as dry weight (mg/100 g of dry tea leaves) and in the form of infusion (mg/100 ml of tea infusions – specific gravities approximately 1.0). Infusions were standardized to 1% (1 g of dry tea leave/100 ml of boiling water). The value identified as n.d. was "not detected" and was provided by the author. This information was included to identify that the component was analyzed but not detected in that food. The lack of values for specific components does not mean that the value is equal to zero, but that the information was not available in the publication. Sources of all information (laboratory or bibliographic reference) were documented in the database. In relation to the information on data quality evaluation, the database includes columns for the total score of each category, for the Quality Index and for the confidence code.

Foods were distributed in the food groups proposed for the LATINFOODS database (FAO, 1995; FAO/LATINFOODS, 2004). Due to flavonoid distribution in foods of plant origin, only the following food groups presented data: B – vegetables and derivatives; C – fruits and derivatives; H – beverages; K – sugar and sweets; N – foods for special diets; Q – infant foods; T – legumes and derivatives.

2.2. Data quality evaluation

The data quality was evaluated through the USDA DQES proposed by Holden et al. (2005) with few modifications, mainly in relation to the distribution conditions of national foods. Also, certain considerations described by Bhagwat et al. (2009) when validating the system were included. The data in each article were evaluated according to five categories: (a) number of samples; (b) sampling plan; (c) sample handling; (d) analytical method; (e) analytical quality control. According to the USDA, these categories represent the major determinants of data quality and this information is essential in order to decide if the data will or will not be included in a database. Scores (0–20 per category) were assigned to the questions, which are specific for each category. The summation of all scores assigned to all categories resulted in the Quality Index (QI) (maximum possible of 100) and a confidence code (CC) was attributed according to the QI range. The CC (A, B, C or D) indicates the relative quality of the data and the reliability of the mean. The CC for flavonoids was assigned as follows (CC, QI value range, meaning of the CC, respectively): A, 75–100, exceptional – the user can have considerable confidence in this value; B, 50–74, above average – the user can have confidence in

Table 1
Distribution of scores in the sampling plan category according to the sample characteristics and probability plan.

Characteristics of sample	Sample probability plan	
	Yes	No
Source of sample (classes)		
I – Bulk from at least 3 different producing regions with cultivar identification; native from producing regions with cultivar identification; manufactured with national distribution	12	10
II – Bulk from 2 main producing regions with cultivar identification; from a commercial warehouse with cultivar identification; native from producing region different than main producer with cultivar identification; manufactured of local distribution	7	3
III – Bulk from one producing region; obtained of one supplier; obtained from experimental lots; experimental cultivar; manufactured without brand identification	1	1
Number of lots or brands		
≥3	6	5
2	4	3
1	2	1
Number of sample units per lot		
≥2	1	1
1	0	0
Number of seasons		
≥2	1	1
1	0	0

Adapted from Holden et al. (2005).

this value, however some problems exist regarding the data on which the value is based; C, 25–49, as average – the user can have less confidence in this value due to limited quantity and/or quality of data; D, <25, below average – there are significant problems with the value related to limited quantity and/or quality of data (Holden et al., 2005; Bhagwat et al., 2009).

2.2.1. Criteria adopted and scores per category

In the sampling plan category, the representativeness of the analytical result of a food or a product in relation to national consumption is considered. The variability of the flavonoid content in foods can be explained by several factors, such as agricultural practices, stress (climate, ultraviolet radiation, light), growing local, seasons, cultivars, storage conditions and others (Aherne and O'Brien, 2002). Therefore, during the evaluation of a data, it is important to consider aspects related to: source of samples, number of lots or brands, number of samples per lots, number of seasons and statistical planning.

This category was most modified to employ the USDA DQES for Brazilian foods. Also, some criteria used in the Brazilian data quality evaluation system for dietary fiber (Menezes et al., 2000) were applied. Foods that are commercialized in bulk or manufactured were evaluated together, and this evaluation considered three classes for sample sources, number of lots or brands, number of sample units per lot and number of seasons. The score distribution in the sampling plan according to the sample characteristics is presented in Table 1. The highest score would be given to a probability plan. In order to facilitate the evaluation of sample sources, three classes were created, which consider the regions of food production and commercialization. Table 1 presents the criteria adopted for each class. For sample classification, national information on the main producing regions of the different foods was considered (IBGE, 2009). Most foods are produced in different regions, sent to warehouses (supply centers) and then distributed to the different states and cities to be commercialized. Therefore, samples from these centers were assigned better score than samples from a specific place.

In the sample handling category, the whole trajectory of the sample is considered, since its acquisition until its analysis. The evaluation of these steps aims to ensure that the stability of the food matrix and nutrient content were preserved. For the questions in this category, the answers received the scores ranged from 0 to 20 (Table 2).

The adequacy of the number of individual samples analyzed is evaluated in the number of samples category. The scores of this category show the reliability of the mean and sample-to-sample variability. In the case of replicates of the same sample, only one sample was considered. Also, a composite sample (resulting from homogenization of many sample units from different areas, lots, brands, etc.) was considered as one sample. It is important to emphasize that in the sampling plan category, the number of lots and/or brands of composite samples were considered. According to the number of samples analyzed, the following scores were assigned, respectively (Holden et al., 2005): 1-1 score; 2-4; 3-7; 4-9; 5-11; 6-13; 7-15; 8-16; 9-17; 10-18; 11-19; >12-20.

The analytical method category considers the performance of the technique for flavonoid analysis and method validation. The same analytical methods for flavonoid analysis selected by Holden et al. (2005) in the elaboration of the UDSA flavonoid database were used. The analysis by high-performance liquid chromatography (HPLC) was adopted as criteria for data inclusion, because it provides good separation and quantification of the compounds in the group of flavonoids (Merken and Beecher, 2000; Holden et al., 2005), whereas results obtained by other methods or by qualitative methods were discarded.

Table 2
Distribution of scores in the sample handling category.

Questions related to sample treatment	Answers	Scores
1. Is homogenization of samples necessary?	Yes (go to 2)	–
	No (skip 2, 3, 4)	10
	Unknown (skip 2, 3, 4)	0
2. Was homogenization performed?	Yes	5
	No	0
3. Was homogenization validated?	Yes	3
	No	0
4. Was information about equipment given?	Yes	2
	No	0.5
5. Was only edible portion analyzed?	Yes	3
	No	0
6. Was moisture information given?	Yes	3
	No	0
7. Were samples stored properly?	Yes	4
	No	0

Adapted from Holden et al. (2005).

Table 3
Distribution of scores in the analytical method category for flavonoids according to critical steps of sample processing, analysis and quantitation (first part).

Questions related to critical steps of sample processing, analysis and quantitation	Scores	
	Yes	No or unknown
1. Were analyte peaks identified by more than one method?	0.5	0
2. If external standardization was used for quantitation, was the purity of standard verified?	0.5	0
3. If internal standardization was used for quantitation, was the standard similar in stability, chemical and spectral properties?	0.5	0
4. Were ≥3 concentrations used for the standard curve?	1.0	0
5. Was the linearity of the standard curve demonstrated?	0.5	0
6. Was the calibration curve coefficient (r) ≥0.99?	0.5	0
7. Was the instrument response checked frequently?	0.5	0
8. Were the samples protected from oxidation (use of TBHQ, BHT, N_2, BHA, etc.)?	0.5	0
9. Was optimization of extraction reported?	1.25	0
10. Were the samples protected from UV light?	0.25	0
11. Was the sample size 5 g (anthocyanidins) or 1 g (other flavonoids)?	0.5	0
12. Were samples hydrolyzed?	0.5	0.5
13. Were losses by hydrolysis minimized?	1.5	0.1
14. If samples were not hydrolyzed, was adequate resolution of peaks demonstrated?	1.5	0.1

Adapted from Holden et al. (2005).

Table 4
Distribution of scores for evaluation of the execution of the flavonoid analytical methodology in the laboratory (second part).

Questions related to method execution	Answers	Scores
1. Was Reference Material (RM) used?	Yes (go to 2)	–
	No (go to 4)	–
	In-house material was used (skip 2)	–
2. What was the range observed according to the Reference Material used?		
Certified Reference Material (CRMs)	Values within accepted range	4
	Values within extended range (±15%)	3
Standard Reference Materials (SRMs)	Values within accepted range	3
	Values within extended range (±15%)	2
In-house material	Go to 3	–
3. What was the range of % of quality control material recoveries?	95–100%	2
	90–110%	1.5
	85–115%	1
	80–120%	0.5
	<80% or >120%, or unknown	0
4. What was the % of difference in results when compared to another laboratory or method?	≤10%	2
	≤15%	1.5
	≤20%	1
	>20%, or unknown	0
5. What was the % of coefficient of variation (CV) observed? (Repeatability studies – precision)	≤10%	2
	≤15%	1.5
	≤20%	1
	>20%, or unknown	0

Adapted from Holden et al. (2005).

The process of scoring the analytical method category considers two parts (Holden et al., 2005). For each part of the analytical method evaluation, the maximum of 10 points can be assigned.

The first part refers to the method itself, where scores are assigned to the main critical steps of sample processing, analysis and quantitation (Table 3). Regarding the pre-treatment of flavonoid hydrolysis, this procedure was evaluated according to the characteristics of each compound studied. The acid hydrolysis has been used by several authors to isolate aglycones (Hertog et al., 1992; Häkkinen et al., 1998); however, this procedure significantly decreases the content of isoflavones in soy, for instance (Genovese and Lajolo, 2001). By subjecting samples of soy to different HCl concentrations and hydrolysis times, the authors observed a significant isoflavone loss and do not indicate this treatment for soy (Genovese and Lajolo, 2001). Using the acid hydrolysis can also degrade other flavonoids such as anthocyanins and catechines (Merken and Beecher, 2000); therefore, each data was evaluated specifically. For instance, in the case of the analysis of catechines in tea, the acid hydrolysis was eliminated (Matsubara and Rodriguez-

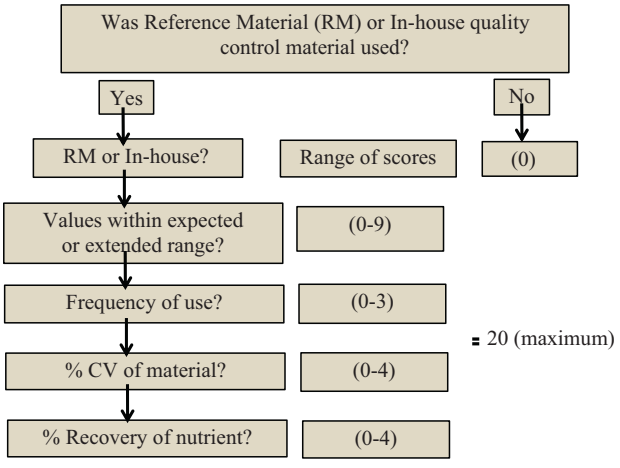

Fig. 1. Range of scores assigned in the sampling plan category.

Amaya, 2006b) and for the analysis of certain compounds in fruits, modifications were made in its procedure (Hoffmann-Ribani et al., 2009). Also, some data were discarded for not being adequate to the conditions necessary to preserve flavonoids during the hydrolysis process. All these details and criteria adopted regarding the methodology used for flavonoid quantification were documented in the compilation forms of each data.

The second part of the process of scoring the analytical method category is related to the evaluation of the execution of the analytical method in the laboratory, where precision and accuracy of the method are considered (Table 4). The precision is evaluated by the percentage of the coefficient of variation (% CV), where the highest score is for CV $\leq$ 10%. Accuracy is determined by the use of Certified Reference Materials (CRMs) or standard Reference Materials (SRMs), where values within accepted or extended range (±15%) for CRMs and SRMs are accepted; or by the use of in-house material (developed for specific nutrients when the CRM is not available), in this case the percentage of recovery of the material is evaluated (highest score is assigned to 95–100% of recovery); or by comparison of data between laboratories, whereas the difference between the value found and that of other laboratory is considered (highest score is assigned to difference $\leq$10%).

The analytical quality control category evaluates accuracy and precision in the day-to-day execution of an analytical method. The category is evaluated by the information on the use frequency of reference material and the coefficient of variation (% CV) in the values obtained. In the case of absence of CRMs or SRMs, the in-house material can be used to estimate daily precision; however, without CRMs or SRMs, the evaluation of accuracy is compromised. The description of the in-house material should be provided (Emons, 2006; Bhagwat et al., 2009), and it should be used daily to each sample batch, and the results should be compared periodically with CRMs or SRMs when available. In this category, the scores for answers are shown in Fig. 1.

Table 5
Example of some information in the Brazilian flavonoid database.

ID[a]	Short food identification	Moisture[b] (g/100 g)	Subclass	Flavonoids	N[c]	Mean[d] (mg/100 g)	SD[e]/variation	CC[f]	Source of data
B1091	Lettuce, curly, raw, from July to December 2001	97.40	Flavonols	Kaempferol	1	n.d[g]		C	Arabbi et al. (2004)
				Quercetin	1	18.4	±0.30	C	
			Flavones	Luteolin	1	0.20	±0.00	C	
				Apigenin	1	n.d.		C	
			Anthocyanidins	Cyanidin	1	n.d.		C	
B1094	Lettuce, purple, raw, from January to June 2002	95.30	Flavonols	Kaempferol	1	n.d.		C	Arabbi et al. (2004)
				Quercetin	1	37.50	±1.90	C	
			Flavona	Luteolin	1	8.80	±0.90	C	
				Apigenin	1	n.d.		C	
			Anthocyanidins	Cyanidin	1	20.80	±1.50	C	
B1095	Lettuce, purple, raw, from July to December 2001	94.20	Flavonols	Kaempferol	1	n.d.		C	Arabbi et al. (2004)
				Quercetin	1	44.90	±2.50	C	
			Flavones	Luteolin	1	3.10	±0.10	C	
				Apigenin	1	n.d.		C	
			Anthocyanidins	Cyanidin	1	19.00	±0.70	C	
C755	Acerola, *in natura*, Waldy, from January to June 2003	91.19	Anthocyanidins	Cyanidin	3	5.53		C	Rosso et al. (2008)
				Pelargonidin	3	0.99		C	
C756	Acerola, *in natura*, Waldy, from January to June 2004	91.19	Anthocyanidins	Cyanidin	3	6.48		C	Rosso et al. (2008)
				Pelargonidin	3	1.16		C	
C759	Acerola, frozen pulp	91.65	Flavonols	Quercetin	1	5.50	±0.20	C	Santos (2005)
			Anthocyanidins	Cyanidin	2	10.30	±0.60	C	
				Pelargonidin	2	1.17	±0.04	C	
H19	Black tea, commercial, infusion 1%	99.60	Flavonols	Kaempferol	4	0.25	±0.01	C	Matsubara and Rodriguez-Amaya (2006a,b)
				Quercetin	3	0.47	±0.03	C	
				Myricetin	3	0.11	±0.01	C	
			Flavan-3-ols	Catechin	4	n.d.		C	
				Epicatechin	1	1.40	±0.10	C	
				Epigallocatechin	1	8.10	±0.40	C	
				Epicatechin gallate	2	3.10	±0.35	C	
				Epigallocatechin gallate	1	6.00	±0.50	C	
				Theaflavin	1	1.90	±0.10	C	
				Theaflavin 3-gallate	2	1.90	±0.10	C	
				Theaflavin 3′-gallate	2	1.15	±0.10	C	
				Theaflavin 3-3′-digallate	1	1.70	±0.10	C	
H30	Soy, extract, natural flavor, "Ades"	86.00	Isoflavones	Glycitein	1	0.87		C	Genovese and Lajolo (2002)
				Genistein	1	4.16		C	
				Daidzein	1	3.26		C	
T136	Soy, flour, without fat	10.50	Isoflavones	Glycitein	6	10.49	5.15–24.01	B	Genovese et al. (2007)
				Genistein	6	113.94	69.10–198.94	B	
				Daidzein	6	65.26	45.20–120.05	B	

[a] ID = identification.
[b] Moisture = g/100 g of edible portion of food.
[c] *N* = number of samples.
[d] Mean = g/100 g of fresh weight of edible portion (expressed as aglycons).
[e] SD = standard deviation.
[f] CC = confidence code.
[g] n.d. = not detected.

3. Results and discussion

Only 22 out of 37 scientific publications and thesis had adequate information to be compiled. Fifteen works were excluded mainly due to the use of spectrophotometric or qualitative methods and data with total flavonoid content. In the compiled works, it was possible to evaluate the quality of 773 flavonoid data, distributed among different subclasses, from 197 foods. The complete Brazilian flavonoid database is available in the master's degree dissertation (Santos, 2009) and Table 5 shows an example of some information present in this database.

When interpreting the values of food flavonoid contents (mg/100 g of fresh weight of edible portion), although the number of samples per food was low, it was possible to observe great variation in the flavonoid content regarding the cultivars of a same food, as evidenced in acerola, beans, strawberry and grape. In acerola, the quantity of cyanidin varied from 5.5 to 38 mg/100 g among three cultivars; in strawberry, for example, the content of pelargonidin, the main flavonoid, varied from 17 to 43 mg/100 g among seven cultivars. Another variation observed was due to the period of harvest, as observed with data of lettuce, chicory, onion and pepper, obtained in different periods; in the case of chicory, the quercetin content was 3.7 mg/100 g from January to June and 25.2 mg/100 g from July to December, while in purple onion the content was 38.3 mg/100 g from July to December and 93.6 mg/100 g from January to June (Santos, 2009). These differences observed are pertinent, since the flavonoid content may vary significantly in plants because of several factors (Aherne and O'Brien, 2002). At the same time, it is important to emphasize that data which did not present reliability were excluded. For example (Table 5), data of black tea present variable number of samples for each flavonoid and they are different from the original paper (Matsubara and Rodriguez-Amaya, 2006a,b), because some data had to be excluded once they presented high coefficient of variation (>15%).

Regarding the evaluation of flavonoid data quality, the confidence code (CC) "B" was attributed to 1% of data (9), which means data with quality above average (the user can have confidence in this value, however some problems exist regarding the data on which the value is based) and "C" was attributed to 99% of the data (764), which means data quality as average (the user can have less confidence in this value due to limited quantity and/or quality of data). The CC "A" (exceptional) and "D" (below average) were not assigned in the Brazilian flavonoid database. The CC "A" was not observed mainly due to the lack of execution of the sampling plan and lack of details in quality control. Even the laboratories that execute the analysis properly were not concerned about providing details on the proceedings regarding quality control in their publications. This category was the only one assigned score zero for the three flavonoid subclasses and hence, did not obtain a CC "A". Several data with any doubt regarding the analytical procedure were discarded (the average score for the category analytical method was 9, Table 6); possibly some of these data could have been assigned CC "D" if they had not been discarded. Another possible explanation for the evaluated publications (22) not being assigned CC "D" may be the reduced number of groups that do the analysis adequately, which may also have influenced the obtainment of the same average score for the analytical method category.

In the USDA flavonoid database (Holden et al., 2005), for 1469 flavonoid data from 225 foods from different countries, the CCs were the following: A – 3% (the user can have considerable confidence in this value); B – 61%; C – 31% and D – 5% (there are significant problems with the value related to limited quantity and/or quality of data). Therefore, most part of data in the USDA database received CC "B" (50–74 scores) and the ones in the Brazilian database, "C" (25–49 scores). This resulting confidence code in the Brazilian database is mainly due to the low score in the number of samples category.

Table 6 shows the average scores obtained in the evaluation of 773 data points, considering that each category can achieve a maximum score of 20. Each average score was classified according to its distribution in the following score ranges: 0–5 considered as below average; 6–10 as average; 11–15 as above average and 16–20 as exceptional.

In the number of samples category, the average score of subclasses was 2 (below average) (Table 6), which was the lowest score in all categories. Most part of data (88%) received score 1, because only one sample was collected for flavonoid analysis in the studies. In the isoflavone subclass, the highest score was assigned to the data of only one work (Genovese et al., 2007) due to data on glycitein, genistein and daidzein of soy from 13 samples. In the USDA flavonoid database, this category received average score of 15 (above average), because of several articles with significant number of samples (Holden et al., 2005). For the most part Brazilian researchers do not aim to produce data for food composition tables; but for research involving control and experimental samples, the number of samples was low in the compiled works. This kind of procedure for number of samples must be modified so the information produced is representative and widely used.

Although the evaluation of the sampling plan category requires only basic information (Table 1), most of the compiled articles did not provide it and/or mention any probability plan for collection of samples, showing that these procedures were not considered. The lowest score for the sampling plan category was 2 (Table 6) (22% of data), which means that samples were obtained from a local supplier or one producing region, or manufactured without brand identification (class III), one lot, and analyzed in only one sample. The highest score (11) (only 16% of data), refers to the samples that fit into classes I and II of source sample (provided from several regions and cultivars, or because they were obtained in a warehouse or they are from native foods obtained in the producing region – Table 1), from a bigger number of lots, analyzed in several samples per lot and obtained in more than one season. Therefore, the compiled data are very distant from an ideal sampling plan that

Table 6
Average scores (minimum and maximum) obtained after data evaluation according to the five categories and flavonoid subclasses in the Brazilian database.

Categories/Subclasses	Number of samples	Sampling plan	Sample handling	Analytical method	Analytical quality control
	Mean (minimum–maximum)				
Flavonols	1 (1–9)	4 (2–11)	20 (17–20)	9 (9)	3 (0–4)
Flavones	1 (1–4)	4 (4–7)	20 (20)	9 (9)	4 (2–4)
Flavanones	1 (1)	4 (4)	20 (20)	9 (9)	3 (2–4)
Flavan-3-ols	1 (1–9)	5 (2–11)	20 (20)	9 (9)	4 (0–4)
Isoflavones	3 (1–20)	10 (6–11)	18 (17–20)	9 (9)	5 (0–8)
Anthocyanidins	2 (1–7)	5 (2–11)	20 (17–20)	9 (9)	3 (0–4)
Average of all subclasses	2 (1–20)	5 (2–11)	20 (17–20)	9 (9)	4 (0–8)

guarantees the sample representativeness for this kind of component (Greenfield and Southgate, 2003; Charrondiere et al., 2009). The average score in this category was 5 (below average), varying from 2 to 11 (Table 6). In the USDA flavonoid database (Holden et al., 2005), the average score was 10 (average), varying from 2 to 14. In this case, the highest score was assigned to data of products from several countries, for example, the flavonoid content of wines from Spain, Germany and United States, increasing the score in this category.

The average score in the sample handling category (Table 6) was the best one among all five categories, varying from 17 to 20 (exceptional). The highest score (20) was assigned to 88% of data. The articles presented enough information regarding the sample handling since its collection until its analysis. Only 12% of data was assigned score 17, due to the lack of information on moisture. The moisture content in food composition tables is essential, once it interferes in the calculation and interpretation of the content of other components, and also allows the comparison of nutritional values with other databases (Greenfield and Southgate, 2003; Charrondiere et al., 2009). In the USDA flavonoid database (Holden et al., 2005), the average score in this category was 17 (12–20). The authors of the works compiled in both databases presented significant information for the category, resulting in a high score.

In the analytical method category, the average score was equal to all subclasses, with score 9 (average) (Table 6). The average score for analytical method was not lower (9), once the mean score for the first part of method evaluation (processing, analysis and quantitation) was high. However, details related to the method validation (accuracy and precision) and use of reference material were not mentioned in most part of the articles and this part received low scores. Similar score (9, varying from 2 to 15) was found in the USDA flavonoid database (Holden et al., 2005). The same problems were identified in both databases, such as lack of certificate or standard reference material for flavonoids, details about variability of the analytical process (% CVs) and about the range of % of quality control material recoveries. Analysts need to be aware of the necessity of describing this information in a detailed way in the publications. In addition, Bhagwat et al. (2009), during the validation of the quality evaluation systems of data on different components, emphasized that the compiler has to have some experience and/or training in chemistry, so they will be able to evaluate this category adequately (for example, what CRMs, SRMs and in-house material are and for what purpose they are used).

The average score obtained in data evaluation in the analytical quality control category was low, equal to 4 (below average), varying from 0 to 8. The subclass that received the highest score was isoflavones, with score 8, from 10% of the total data. This highest score was assigned only to data from one single work, which provided adequate information on CV ($\leq$5%) and percentage of recovery varying from 95 to 100% (Genovese and Lajolo, 2002). Among all categories, this was the only one that received score zero (in 12% of compiled data, in the subclasses flavonols, flavan-3-ols, isoflavones and anthocyanidins). Similar results were observed in the USDA database (Holden et al., 2005), with average score 5, varying from 0 to 17. Therefore, data on the analytical quality control category presented significant deficiency in both databases.

In order to evaluate the accuracy and precision in the day-to-day execution of an analytical method, it is necessary to know, at least, the use frequency of reference material and the coefficient of variation (% CV) in the values obtained. However, this information is rarely provided by publications, even the use of in-house material is seldom mentioned, which turns this category the most neglected by analysts. Bhagwat et al. (2009) warned about a great difficult of compilers in evaluating data in this category. Adequate procedures in these two categories, analytical method and analytical quality control, have been widely discussed and divulged by the USDA (Bhagwat et al., 2009; Haytowitz et al., 2009), EuroFIRr project (Castanheira et al., 2007; Westenbrink et al., 2009) and others (Greenfield and Southgate, 2003; Charrondiere et al., 2009), aiming to warn the analysts about the necessity of using good laboratory practices and the improvement of the data analytical quality.

In general, the process of quality evaluation of flavonoid national data allowed the identification of categories that received low average scores and their deficiencies, respectively: number of samples (2), for the low number of samples; sampling plan (5), due to the lack of probability plan; analytical quality control (4), due to the lack of description of the method daily performed in the laboratory. The second part of analytical method category is also deficient because of the lack of description and execution of the analytical method. Therefore, the four categories are critical in relation to national data. Analysts and researchers need to be aware of the number and planning of samples to be analyzed and really should document all the process of methodology execution and analytical quality control.

Dissemination of flavonoid Brazilian database – flavonoids are bioactive compounds present in foods of vegetable origin. Due to their anti-inflammatory, antioxidative and antimicrobial properties, certain flavonoids can be associated with cardioprotective and/or anticarcinogenic effects (Kris-Etherton et al., 2004; Gry et al., 2007; Denny and Butriss, 2007). The next step is to introduce the flavonoid Brazilian database in the Brazilian Food Composition Database-USP (TBCA-USP) (http://www.fcf.usp.br/tabela) (USP, 1998), as previously done with other compounds (Menezes et al., 2002, 2009). Through the dissemination of the main flavonoid compounds of 197 Brazilian foods, the user will have an important tool to help on decreasing the risk for non-transmissible chronic diseases (NTCD).

4. Conclusion

The compilation of 22 scientific publications resulted in 773 flavonoid data, distributed in six subclasses (flavonols, flavones, isoflavones, flavanones, flavan-3-ols, and anthocyanidins), from 197 Brazilian foods. The quality evaluation of the flavonoid data resulted in 764 data (99%) with CC "C" (as average) and 9 data (1%) with CC "B" (above average).

Acknowledgements

The authors acknowledge the financial support from USP and CNPq and the participation by BRASILFOODS. Special acknowledgement to Ms. Joanne M. Holden of Nutrient Data Laboratory, ARS/USDA for making all material of the USDA DQES available and for discussing the modifications necessary for its use in Brazilian foods.

References

Aherne, S.A., O'Brien, N.M., 2002. Dietary flavonols: chemistry, food content, and metabolism. Nutrition 18 (1), 75–81.
Arabbi, P.R., Genovese, M.I., Lajolo, F.M., 2004. Flavonoids in vegetable foods commonly consumed in Brazil and estimated ingestion by the Brazilian population. Journal of Agricultural and Food Chemistry 52, 1124–1131.
Bhagwat, S.A., Patterson, K.Y., Holden, J.M., 2009. Validation study of the USDA's Data Quality Evaluation System. Journal of Food Composition and Analysis 22, 366–372.
Burlingame, B., Charrondiere, R., Mouille, B., 2009. Food composition is fundamental to the cross-cutting initiative on biodiversity for food and nutrition. Journal of Food Composition and Analysis 22, 361–365.
Castanheira, I., Robb, P., Owen, L., Boer, H., Schmitt, J., Ent, H., Calhau, M.A., 2007. A proposal to demonstrate a hamonized quality approach to analytical data

production by EuroFIR. Journal of Food Composition and Analysis 20 (8), 725–732.
Charrondiere, U.R., Burlingame, B., Berman, S., Elmadfa, I., 2009. Food Composition Study Guide, vol. 1/2. INFOODS, FAO, Rome Retrieved November, from http://www.fao.org/infoods/publications_en.stm.
Denny, A., Butriss, J., 2007. Plants food and health: focus on plant bioactives. EuroFir, British Nutrition Foundation, Norwich, 64 pp. (synthesis reports, no. 4). Retrieved 2009 from http://www.eurofir.net/temp/PLANTspFOODSspANDspHEALTHspFOCUSspONspPLANTspBIOACTIVEShs1hs.pdf.
Emons, H., 2006. The 'RM family' – identification of all its members. Accreditation and Quality Assurance 10, 690–691.
Food and Agriculture Organization (FAO), 1995. INFORME del taller CTPD sobre production y manejo de dados de composición quimica de alimentos en nutrición de América Latina. FAO/Oficina Regional para la América Latina y el Caribe (NUT-60), Santiago.
Food and Agriculture Organization/Red Latinoamericana de Datos de Composición de Alimentos (FAO/LATINFOODS), 2004. Segunda Conferência Eletrônica -Avaliação da qualidade dos dados para bases de dados e tabelas de composição química de alimentos. 11 a 29 de novembro de 2004. Retrieved 2009-01-16 from http://www.rlc.fao.org/foro/latinfoods/.
Genovese, M.I., Barbosa, A.C.L., Pinto, M.S., Lajolo, F.M., 2007. Commercial soy protein ingredients as isoflavone sources for functional foods. Plant Foods Human Nutrition 62, 53–58.
Genovese, M.I., Lajolo, F.M., 2001. Determination of isoflavones in soy products. Ciência Tecnologia Alimentos 21, 86–93.
Genovese, M.I., Lajolo, F.M., 2002. Isoflavones in soy-based foods consumed in Brazil: levels, distribution, and estimated intake. Journal of Agricultural and Food Chemistry 50, 5987–5993.
Greenfield, H., Southgate, D.A.T., 2003. Food Composition Data: Production, Management and Use, second ed. FAO, Rome Retrieved from ftp://ftp.fao.org/docrep/fao/008/y4705e/y4705e00.pdf.
Gry, J., Black, L., Eriksen, F.D., Pilegaard, K., Plumb, J., Rhodes, M., Sheehan, D., Kiely, M., Kroon, P.A., 2007. EuroFIR-BASIS: a combined composition and biological activity database for bioactive compounds in plant-based foods. Trends in Food Science & Technology 18, 434–444.
Häkkinen, S.H., Kärenlampi, S.O., Heinonen, M.I., Mykkänen, H.M., Törrönen, A.R., 1998. HPLC method for screening of flavonoids and phenolic acids in berries. Journal of the Science of Food and Agriculture 77, 543–551.
Haytowitz, D.B., Lemar, L.E., Pehrsson, P.R., 2009. USDA's Nutrient Databank System – a tool for handling data from diverse sources. Journal of Food Composition and Analysis 22, 433–441.
Hertog, M.G.L., Hollman, P.C.H., Katan, M.B., 1992. Content of potentially anticarcinogenic of flavonoids of 28 vegetables and 9 fruits commonly consumed in the Netherlands. Journal of Agricultural and Food Chemistry 40 (12), 2379–2383.
Hoffmann-Ribani, R., Huber, L.S., Rodriguez-Amaya, D., 2009. Flavonols in fresh and processed Brazilian fruits. Journal of Food Composition and Analysis 22, 263–268.
Holden, J.M., Bhagwat, S.A., Haytowitz, D.B., Gebhardt, S.E., Dwyer, J.T., Peterson, J., Beecher, G.R., Eldridge, A.L., Balentine, D., 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., Schubert, A., 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 (3), 285–299.
Instituto Brasileiro de Geografia e Estatística (IBGE). Mapa do Site. Especiais! Os Estados Brasileiros (7 a 12). Retrieved May 1, 2009 from http://www.ibge.gov.br/7a12/especiais/estados/estados_brasil.html.
International Network of Food Data Systems/Food and Agriculture Organization (INFOODS/FAO). Tagnames for food components. Retrieved May 20, 2009 from http://www.fao.org/infoods/tagnames_en.stm.
Kris-Etherton, P.M., Lefevre, M., Beecher, B.R., Gross, M.D., Kenn, C.L., Etherton, T.D., 2004. Bioactive compounds in nutrition and health-research methodologies for establishing biological function: the antioxidant and anti-inflammatory effects of flavonoids on atherosclerosis. Annual Review of Nutrition 24, 511–538.
Mangels, A.R., Holden, J.M., Beecher, G.R., Forman, M.R., Lanza, E., 1993. Carotenoid content of fruits and vegetables: an evaluation of analytical data. Journal of the American Dietetic Association 93 (3), 284–296.
Matsubara, S., Rodriguez-Amaya, D.B., 2006a. Conteúdo de miricetina, quercetina e kaempeferol em chás comercializados no Brasil. Ciência Tecnologia Alimentos 26 (2), 380–385.
Matsubara, S., Rodriguez-Amaya, D.B., 2006b. Teores de catequinas e teaflavinas em chás comercializados no Brasil. Ciência Tecnologia Alimentos 26 (2), 401–407.
Menezes, E.W., 2009. 7th International Food Data Conference: food composition and biodiversity. Journal of Food Composition and Analysis 22 (5), 359–360.
Menezes, E.W., Caruso, L., Lajolo, F.M., 2000. An application of criteria to evaluate quality of dietary fibre data in Brazilian foods. Journal of Food Composition and Analysis 13 (4), 455–473.
Menezes, E.W., Giuntini, E.B., Dan, M.C.T., Lajolo, F.M., 2009. New information on carbohydrates in the Brazilian Food Composition Database. Journal of Food Composition and Analysis 22, 446–452.
Menezes, E.W., Gonçalves, F.A.R., Giuntini, E.B., Lajolo, F.M., 2002. Brazilian food composition database: Internet dissemination and others recent developments. Journal of Food Composition and Analysis 15 (4), 453–464.
Menezes, E.W., Ratto, A.T., Giuntini, E.B., Lajolo, F.M., 2005. composição de alimentos: compilação e uniformização de estruturas para intercâmbio de dados. Brazilian Journal of Food Technology 8 (1), 25–33.
Merken, H.M., Beecher, G.R., 2000. Measurement of food flavonoids by high-performance liquid chromatography: a review. Journal of Agricultural and Food Chemistry 48 (3), 577–599.
Rosso, V.V., Hillbrand, S., Montilla, E.C., Bobbio, F.O., Winterhalter, P., Mercadante, A.Z., 2008. Determination of anthocyanins from acerola (*Malpighia emarginata* DC.) and açaí (*Euterpe oleracea* Mart.) by HPLC–PDA–MS/MS. Journal of Food Composition and Analysis 21, 291–299.
Santos, N.C., 2009. Tabela Brasileira de Composição de Alimentos (TBCA-USP): Banco de dados de composição de bioativos. Master's degree thesis. Post Graduation Course PRONUT – FCF/FEA/FSP-USP, University of São Paulo, Brazil.
Santos, R.J., 2005. Efeito do armazenamento de polpas de frutas congeladas sobre os teores de flavonóides, vitamina C e atividade antioxidante. Master's degree thesis. Post Graduation Course Food Science – FCF-USP, University of São Paulo, Brazil.
Toledo, A., Burlingame, B., 2006. Biodiversity and nutrition: a common path toward food security and sustainable development. Journal of Food Composition and Analysis 19, 477–483.
Truswell, A.S., Bateson, D.J., Madaglio, K.C., Pennington, J.A.T., Rand, W.M., Klensin, J.C., 1991. INFOODS Guidelines for describing foods: a systematic approach to describing foods to facilitate international exchange of food composition data. Journal of Food Composition and Analysis 4, 18–38.
Universidade de São Paulo (USP), Faculdade de Ciências Farmacêuticas, Departamento de Alimentos e Nutrição Experimental, 1998. Tabela Brasileira de Composição de Alimentos-USP. Version 5.0. Retrieved October 10, 2009 from http://www.fcf.usp.br/tabela.
U.S. Department of Agriculture, Agricultural Research Service, 2007. USDA Database for the flavonoid content of selected foods. Release 2.1. Nutrient Data Laboratory. Retrieved May 28, 2010 from http://ars.usda.gov/nutrient data.
U.S. Department of Agriculture, Agricultural Research Service, 2008. USDA Database for the isoflavone content of selected foods. Release 2.0. Nutrient Data Laboratory. Retrieved May 28, 2010 from http://ars.usda.gov/nutrient data.
Westenbrink, S., Oseredczuk, M., Castanheira, I., Roe, M., 2009. Food composition databases: the EuroFIR approach to develop tools to assure the quality of the data compilation process. Food Chemistry 113, 759–767.