Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Food Composition and Analysis 18 (2005) 829–844
JOURNAL OF FOOD COMPOSITION AND ANALYSIS
www.elsevier.com/locate/jfca

Original Article

Development of a database of critically evaluated flavonoids data: application of USDA's data quality evaluation system

Joanne M. Holden[a,*], Seema A. Bhagwat[a], David B. Haytowitz[a], Susan E. Gebhardt[a], Johanna T. Dwyer[b], Julia Peterson[b], Gary R. Beecher[a], Alison L. Eldridge[c], Douglas Balentine[d]

[a]*Nutrient Data Laboratory, Beltsville Human Nutrition Research Center, Bldg. 005, Rm. 201, BARC West, 10300 Baltimore Blvd, ARS/USDA, Beltsville, MD 20705-2350, USA*
[b]*Tufts University Friedman School of Nutrition Science & Policy, Boston, MA, USA*
[c]*Bell Institute of Health and Nutrition, General Mills, Minneapolis, MN, USA*
[d]*Unilever BestFoods North America, Englewood Cliffs, NJ, USA*

Received 12 January 2004; received in revised form 11 June 2004; accepted 14 July 2004

Abstract

The USDA Special Interest Database for flavonoid content of selected foods contains 26 most abundant compounds within 5 predominant subclasses of flavonoids–flavonols, flavones, flavanones, flavan-3-ols, and anthocyanidins. All the data were evaluated for 5 quality evaluation categories (sampling plan, sample handling, analytical method, analytical quality control and number of samples), using the data quality evaluation system developed by the USDA scientists. Confidence Codes (A–through D) were then assigned to every value. The database contains acceptable values for 225 selected foods. Only 97 sources out of approximately 475 collected included acceptable analytical data. The overall quality of data was good with 64% of the observations receiving A or B confidence codes; the flavan-3-ols subclass received better ratings than other subclasses. While this is the first comprehensive database for flavonoids in foods, the majority of data came from Europe and countries other than the US. Due to the observed variability in the values it will be important to have data for US foods. The evaluation of data quality helps set priorities and further

*Corresponding author. Fax: +1-301-504-0632.
E-mail address: holdenj@ba.ars.usda.gov (J.M. Holden).

0889-1575/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.jfca.2004.07.002

identifies the foods to be analyzed as well as areas to improve data quality. Furthermore, release of data quality confidence codes with data provides necessary information to investigators to assess the impact of flavonoid intake on risk of various chronic diseases.
Published by Elsevier Inc.

Keywords: Flavonoids; Data quality evaluation; Confidence codes

1. Introduction

Flavonoids, the major group of phenolic compounds, are secondary plant metabolites. More than 5000 different flavonoid compounds have been identified. Most of the flavonoids are present in nature as glycosides except for flavan-3-ols (catechins and theaflavins) that are present in either free form or as gallic acid esters (e.g., in tea). The glycosidic linkages appear to be important for the absorption of these compounds (Hollman et al., 1999). However, once absorbed, the bioactivity depends on the forms and the polarity of the compounds circulating in vivo (Rice-Evans et al., 2000).

Recent interest by the scientific community in the types and levels of flavonoids in foods centers on their antioxidative (Rice-Evans et al., 1995), antimicrobial (Rauha et al., 2000), and anti-inflammatory (Formica and Regelson, 1995; Middleton Jr., 1996) properties, which may have possible cardioprotective and/or anticarcinogenic effects associated with certain of them. Hertog (1996) observed reduced risk of coronary heart disease (CHD) in both the Seven Countries Study and the Zutphen Elderly Study among those participants with high intakes (>29.9 mg/day) of flavonoids, but did not observe any effect on cancer risk. Le Marchand et al. (2000) observed an inverse association between quercetin (flavonol) intake and risk of lung cancer. Onions and apples were the major contributors of flavonoids in that study. Hertog et al. (1993a) estimated an average intake of potentially anticarcinogenic flavonoids, flavonols (quercetin, kaempferol, myricetin), and flavones (apigenin and luteolin), of 23 mg/day in aglycone forms in the Dutch population. Justesen et al. (1997) estimated a similar intake of 26 mg/day for the same flavonoids in the Danish population. Their estimates were considerably lower than the average intake of 1 g/day for total flavonoids estimated by Kühnau (1976), even after adjustments for 170 mg as glycoside residues of the three flavonoid subclasses: flavonols, flavones and flavanones, which would equal 115 mg as aglycones. The wide discrepancies may have been due in part, to differences in the compounds included, inappropriate methodologies used to estimate intakes and to analyze foods and the use of values for whole foods instead of edible parts only (Hertog et al., 1993a).

Reliable values for flavonoids in foods are needed to test the possible relationships of flavonoids with cardiovascular disease and cancer risk reduction. To address these needs the United States Department of Agriculture (USDA) has developed a Special Interest Database of critically evaluated analytical data on selected compounds within five subclasses of dietary flavonoids. This database complements the earlier release on the isoflavonoids, another subclass of flavonoids. A Special Interest Database or Table is a focused compilation of acceptable existing data for a specific component or a class of components usually containing a limited list of 150–250 foods. To

determine whether the data are acceptable, they are critically evaluated, rated and then compiled to produce the initial database. This database includes quantitative values and other statistical information and the related quality codes for each food, component and source document. The quantity and quality of existing data and gaps identified are useful in setting priorities for further research. We decided to include five subclasses of flavonoids that comprise most of the monomeric dietary flavonoids: flavonols, flavones, flavanones, flavan-3-ols and anthocyanidins. Twenty-six of the most commonly occurring flavonoids in foods over these five subclasses were included in the database. Isoflavones were not included in this database because a Special Interest Table for isoflavones had already been released in 1999 on the Nutrient Data Laboratory (NDL), USDA website (www.nal.usda.gov/fnic/foodcomp). USDA scientists are also in the process of compiling a database for polymeric flavonoids—proanthocyanidins, which range from dimers to polymers. The authors of the flavonoids database represent the Nutrient Data Laboratory of the Agricultural Research Service, USDA, collaborating with academic and food industry partners.

2. Materials and methods

2.1. Data collection

The general approach to preparing a Special Interest Database includes the collection and evaluation of the existing data from published literature as well as available unpublished analytical data. For flavonoids, key words for different compounds of interest (flavonoids, quercetin, catechin, etc.) were used to conduct literature searches in various scientific databases through the Current Awareness Literature Service (CALS). It retrieved citations from agriculture, biology, and environmental sciences; clinical medicine, life sciences; physical, chemical and earth sciences and social and behavioral sciences. In addition, taxonomic names, genus and species for tea and citrus fruits (Camellia and Citrus, respectively) were used to search the Food Science and Technology Abstracts (FSTA) database. The abstracts were reviewed and relevant articles were retrieved. Articles containing analytical data were separated from those containing qualitative information. Simultaneously, the investigators reviewed and discussed the merits of available analytical procedures and defined acceptable methods. Methods judged to provide good separation of flavonoid compounds were column chromatography and high-performance liquid chromatography (HPLC). Articles that contained data generated by thin layer chromatography (TLC) or paper chromatography, radioimmunoassay (RIA), pH differential methods or only spectrophotometric quantitation were rejected due to the lack of specificity of resolution of the compounds. Furthermore, since the objective was to collect values for specific flavonoid compounds, values for total flavonoids or totals by subclass of flavonoids were not included. If articles reported data on dry weight basis and moisture contents were not given, efforts were made to get either fresh weight values or moisture contents from the authors. If these efforts were not successful data from these articles were not included in the database. For certain cases it seemed reasonable to use the moisture contents from the Standard Reference (SR) database of the USDA. Other details of the inclusion or exclusion criteria for articles are given in Table 1.

Table 1
Summary of Literature review

Number of articles	Explanation of the number and disposition of articles	
475	Total number of articles retrieved by the literature search (approximately 275 on fruits, vegetables, and beverages other than citrus fruits and tea and 200 on citrus fruits and tea)	
378	Total number of articles excluded for various reasons	
	No.	Reason for exclusion
	80	Development of analytical methods for identification only, no analytical data
	37	Data on compounds not used in this database or data given as totals of a subclass instead of as individual compounds of the subclass
	40	Insufficient documentation for evaluation, data reported on dry weight basis without information on moisture content or unable to evaluate because of the non-English language
	16	General review articles, no analytical data
	14	Articles on antioxidant activity of the flavonoids, no analytical data
	5	Articles on bioavailability of the flavonoids, no analytical data
	158	Articles on citrus fruits and tea weeded out for various reasons mentioned above
	28	Rejected after critical review for unacceptable analytical procedures or methods, old data, or unusual food
97	Total number of articles that contributed data to this database	

2.2. Data quality evaluation

The quality of data for each food and compound was evaluated using procedures defined by scientists at the NDL as part of the development of a multinutrient data evaluation module, National Nutrient Databank System (Holden et al., 2002). These procedures were based on general criteria described earlier (Holden et al., 1987; Mangels et al., 1993) with some modifications including critical methodological steps pertinent to flavonoid analysis. The data in each article were evaluated according to five categories: sampling plan, sample handling, number of samples, analytical method and analytical quality control. These categories had been defined by Agricultural Research Service (ARS)/USDA scientists as those that influence the quality of data points to be considered for inclusion in food composition databases representative of a particular food supply. Within each category, specific questions were defined to describe the critical steps necessary for achieving accurate and representative data. The information presented in each publication/report was rated, considering specific criteria for each category, on a scale of 0–20 points per category. The ratings for the five categories for each food and compound were summed to yield a Quality Index (QI) with the maximum possible score equal to 100 points. A Confidence Code (CC) was derived from the QI to indicate the relative quality of the data and the

reliability of a given mean and is provided with the quantitative data for a given food and component. The CC was assigned as follows:

QI	CC
75–100	A (exceptional)
74–50	B (above average)
49–25	C (average)
<25	D (below average)

To rate "Sampling Plan" category the representativeness of the set of sample units must be defined relative to the primary objective for developing the database. For example, USDAs databases are widely used for monitoring the effects of dietary intake of specific components, for nutrition research, and for development of food policy to address the public health issues of the US population. Therefore, the rating for sampling plan reflects the representativeness of food sample units collected for analysis in a specific study. Detailed information about the collection of the sample units, including product description, site of procurement (retail or point of production), location, season, etc., improves the rating because points can be awarded only when details are known. Analysis of samples composited from units obtained from different areas yields more representative data, although information about unit-to-unit variability is lost. The USDA module takes into consideration the number of regions, number of cities in each region, number of locations in each city, number of lots at each location and number of seasons in which the samples were collected to generate the ratings for the sampling plan. Due to the international nature of the data in the flavonoids database the prior criteria for sampling plan based on nationwide sampling were modified. Since European sources dominated in the collection of publications for flavonoids data for this database, number of countries was substituted for number of regions in the rating process, particularly at the point of data aggregation.

The review of *Sample Handling* procedures assessed all the information on handling sample units from the time of procurement to the point of analysis. This category includes questions about critical steps for storage, homogenization (if necessary) and its validation, part of the food analyzed (whole or edible part only) and moisture content. In addition, procedures for food preparation are also evaluated.

The rating process for *Analytical Method* reviewed the validity of each reported analytical method as well as the application of the method by an analyst. As a result the analytical method rating process now has two facets: validation of the method itself (processing of samples, separation/digestion, analysis and quantitation method) and quality control of the method as performed by the laboratory (accuracy and precision). The review of method validation focused on the critical steps in the processing of samples in every analytical method. The critical steps for processing, analysis and quantitation in the flavonoids method were defined and reviewed by analytical experts working in this field. The use of hydrolysis is a critical step. Anthocyanidin, flavonol and flavone glycosides can be hydrolyzed to aglycones, but flavanones, as well as flavan-3-ols are very unstable in an acidic medium and are destroyed easily. Therefore, flavanones and flavan-3-ols should be extracted neatly, without hydrolysis. The amount of sample analyzed is also important. In addition, the laboratory analyst must also demonstrate his/her ability to obtain accurate and precise results. Evidence that Certified Reference Materials (CRMs) or Standard

Critical steps and points for rating
General analytical methods for flavonoids (validation of method)

Steps for sample processing: Questions	Points	
	Yes	No or Unknown
Are the samples protected from oxidation (use of TBHQ, BHT, N_2, BHA etc.)	0.5	0
Are the samples protected from ultraviolet light?	0.25	0
Was optimization of extraction reported?	1.25	0
Was the sample size 5g (anthocyanidins) or 1g (other flavonoids)	0.5	0
Were samples hydrolyzed?	0.1	0.1
Were losses by hydrolysis minimized?	1.5	0.1
Was adequate resolution of peaks demonstrated if samples were not hydrolyzed?	1.5	0.1

Steps for analysis and quantitation: Questions	Points	
	Yes	No or Unknown
Were analyte peaks identified by one method only?	1	0
Were analyte peaks identified by more than one method? (e.g. retention time, mass spectrometry, nuclear magnetic resonance etc.)	2	0
If external standardization was used for quantitation, was the purity of standard verified?	0.5	0
If internal standardization was used for quantitation, was the standard similar in stability, chemical and spectral properties?	0.5	0
Were ≥3 concentrations of the standard bracketing sample conc. used for the standard curve?	0.5	0
Was the linearity of the standard curve demonstrated?	0.5	0
Was the calibration curve coefficient (r) ≥0.99?	0.5	0
Was the instrument response checked frequently?	0.5	0

Fig. 1. Critical steps and points for rating general analytical methods for flavonoids (validation of method).

Reference Materials (SRMs) were used and that reported values were within the certified range leads to the highest rating for the validation of the method. Information on the limits of detection (LOD) and/or limits of quantitation (LOQ), recovery studies, and the standard deviations or coefficients of variations (%CV) is needed to assess analytical precision. The absence of this information may result in losing some rating points for the analytical method. These critical procedures and related points are shown in Figs. 1 and 2.

Questions under ''Analytical quality control'' assessed procedures used by analysts to assure day-to-day analytical accuracy and precision. This area is in addition to the method validation

Critical steps and points for rating laboratory performance of the flavonoids analytical method (Quality control of the method)

A. Use of reference material:

	Certified	Reference /Consensus
Values within accepted range	9	7
Values within extended range (±15%)	5	3

B. Recoveries of standards:	% Recovery	Rating
	95%-100%	4
	90%-110%	3
	85%-115%	2
	80%-120%	1
	<80%->120%	0

C. Comparison with another Laboratory/Method:

	% Difference	Rating
	≤10%	4
	≤15%	3
	≤20%	2
	>20%	0

D. Repeatability studies (precision)

	%CV	Rating
	≤10%	3
	≤15%	2
	≤20%	1
	>20%	0

Fig. 2. Critical steps and points for rating laboratory performance of the flavonoids analytical method (quality control of the method).

criteria stated above. In the absence of CRM or SRM in-house reference/quality control materials can provide estimates for day-to-day precision. However, without CRMs or SRMs limited assessment of the accuracy of measurements can be accomplished. The in-house reference or quality control materials should be run with every batch of samples every day and linked to results of CRMs or SRMs when available. This category was judged by the information provided on frequency of use of such materials as well as the coefficient of variation (%CV) in the values obtained.

The objective in developing the rating scale for the ''Number of Samples'' was to consider the adequacy of the number of samples (individual or composite samples) analyzed in order to obtain a reliable estimate of the mean and a robust estimate of sample-to-sample variability. Documentations concerning the number of samples analyzed were reviewed. For example, ''*n*'' should represent the number of discrete analytical samples analyzed, not the replicate analyses of aliquots from the same homogenate. A composite (of many sample units from different areas, brands, etc.) was considered as one sample no matter how many sample units were combined. However, analysis of a composite may have received additional points for the ''sampling plan'' category.

Re-evaluation of ratings at data aggregation: When the data for similar foods from different studies were aggregated the "Sampling Plan" for the data set was re-evaluated to assess its representativeness of the area of interest (e.g., the US food supply). This is due to the inclusion of data for sample units originating in different regions (or countries in this database). Similarly, ratings for the number of samples category were recalculated based on the total number of samples in the aggregate. The ratings for the other categories, sample handling, analytical method and analytical quality control were averaged for the aggregated data.

2.3. Database description

To date most of the available analytical procedures use a hydrolysis step to convert glycosides into aglycones and, thus, the results are reported as aglycones. Therefore, it was decided to report values for flavonoids in this database as aglycones. When individual glycosides were reported the values were converted into aglycone forms using conversion factors based on the molecular weights of the respective compounds. Some examples of conversion factors are given in Table 2. Data reported on dry weight basis were converted into fresh weight basis if the moisture contents data were available or by using moisture contents from the SR when it seemed reasonable. The values for gallic acid esters of catechins and theaflavins present in tea were not converted into free catechins or theaflavins but were reported as such.

Considerable data were available for different kinds of teas, a majority of which were provided by Unilever Bestfoods, North America (unpublished data). The practice of making tea infusions varies in different countries and according to individual preferences due to the amount, blend, and kind (loose leaf, bag, particle size) of tea used, and to brewing times. Therefore, it is difficult to compare flavonoid data for brewed teas obtained from different sources. Since Arts et al. (2000a), and Hertog et al. (1993b) demonstrated that a majority of tea flavonoids are extracted into the infusion after only short brewing times and do not increase substantially with extended brewing times, adjustment for brewing time was not undertaken. However, catechin and flavonol contents in tea infusions increased almost linearly with the amount (weight) of tealeaves used for brewing. Therefore, all infusion values were standardized to a 1% infusion. These values were calculated using the weight of the tea bag (or loose tea leaves) used to make infusions. Values for tea are given as mg/100 g (100 ml) of tea infusions (as consumed). A separate table of dried teas was also provided (mg/100 g of dry tea) with the database for users who wish to calculate different infusion strengths. Values for beverages were adjusted by their respective specific gravities and reported

Table 2
Examples of conversion factors (glycosides to aglycones)*

Glycoside (Name) mol. wt.	Aglycone (Name) mol. wt.	Conversion factor
(Cyanidin-3-glucoside) 420	(Cyanidin) 288	0.6400
(Quercetin-3-glucoside) 464	(Quercetin) 302	0.6509
(Apigenin-7-glucoside) 432	(Apigenin) 270	0.6250
(Hesperidin) (Hesperetin-7-rutinoside) 611	(Hesperetin) 302	0.4943

*Gallate esters of catechin were reported without any conversions.

"as served". These adjustments were needed only for fruit juices since the specific gravities for tea and wine each are approximately 1.0.

"Trace" values were calculated by multiplying the Limit of Quantitation (LOQ) by 0.71, the factor derived statistically and reported by Mangels et al. (1993). If the LOQ was not available, the value was reported as zero. The LOQ is defined as the lowest point at which the method can quantify the amount of a component in the sample. It represents some multiple of the analytic detection limit. For the HPLC technique it is generally assumed to be 2.5 times the Limit of Detection (LOD). The LOD is the lowest value at which a compound can be detected by the method, but it is an amount that cannot be quantified. It is indicated by a very small peak above the noise level (base line) in the chromatogram. All the values in the flavonoids database were reported as mg/100 g of fresh weight of edible portion of food.

A zero value reported in the database indicates the actual determination of a zero concentration, that is, the absence of any noticeable peak at the baseline of the chromatogram. The lack of a value for a particular flavonoid in a food does not imply a zero value, but only that data were unavailable. The table of analytical values contained values for only those compounds and foods that were available in the literature at the time of the survey; the absence of data for specific foods and compounds does not mean that other classes of compounds are not present in that particular food. For example, while red or black grapes may contain anthocyanidins, no values for anthocyanidins were listed in the table, because acceptable data for these compounds were not available. A precise analytical method for the resolution of individual anthocyanidin peaks is not widely available and therefore limited data have been generated by interested laboratories. As mentioned earlier, values for "total anthocyanidins" analyzed by the pH differential method and reported as equivalent of the standard used for quantitation, were not used in the database. For few foods serious discrepancies were observed if the "total anthocyanidins" values were calculated by summing values reported for individual anthocyanidins by HPLC analysis for the same food. For example, the total anthocyanidins value in sweet cherries, Bing variety, calculated by summing individual anthocyanidins was 117.42 mg/100 g (cyanidin, 111.43, pelargonidin, 0.84, and peonidin, 5.15 mg/100 g) (Gao and Mazza, 1995), while the value for "total anthocyanidins" reported as malvin was 1.88 for the same variety (Heinonen et al., 1998).

During the preliminary review of sources and related documentation it was necessary to make some decisions regarding the structure of the database before data entry could begin. As mentioned earlier, the five subclasses: flavonols, flavones, flavanones, flavan-3-ols and anthocyanidins, which include most of the dietary flavonoids were chosen for inclusion in the database. This database focused only on the most commonly occurring monomeric compounds from each subclass in foods, which made a total of 26 compounds. Thus, the subclasses and compounds included the following:

Flavonols: Quercetin, Kaempferol, Myricetin, Isorhamnetin.

Flavones: Apigenin, Luteolin.

Flavanones: Hesperetin, Naringenin, Eriodictyol.

Flavan-3-ols: (+)-catechin, (+)-Gallocatechin, (−)-Epicatechin, (−)-Epigallocatechin, (−)-Epicatechin 3-gallate, (−)-Epigallocatechin 3-gallate, Theaflavin, Theaflavin 3-gallate, Theaflavin 3′-gallate, Theaflavin 3,3′ digallate, Thearubigins.

Anthocyanidins: Cyanidin, Delphinidin, Malvidin, Pelargonidin, Peonidin, Petunidin.

Data entry: Articles which contained analytical data were scrutinized for the analytical methods used and only the ones that used column chromatography or HPLC were selected for data entry. Information related to food description, analytical values, sample handling, analytical method, sample planning, analytical quality control and number of samples was entered in Excel files. Analytical values were converted into aglycone forms or adjusted for specific gravity whenever needed. Points were assigned to all the critical steps within each of the evaluation categories and ratings were calculated for each category by summing the individual points for the respective category.

Following the assignment of ratings to each value for each food and compound, foods were assigned nutrient databank (NDB) identification numbers using detailed food description information and procedures defined for the SR. Because the data came from various sources, both in the United States and other countries, there were a number of foods that are not included in the SR. In those cases, the authors assigned a temporary NDB number beginning with "99". Those numbers are not unique between databases, as they may have been used in other Special Interest Databases produced by NDL, but they are unique within the flavonoids database. Data values were aggregated by flavonoid and NDB number. Subsequently, the mean value (mg/100 g), standard error of the mean (SEM), minimum (Min.) and maximum (Max.) values were calculated for each food and flavonoid compound. Most sources reported mean values, and in some cases, the number of samples and the standard error were also reported. Mean values were weighted by their respective number of samples. The weighted means, in turn, were used to calculate the standard errors based on the total number of samples in each aggregated food.

The flavonoid values for the selected foods along with the confidence code and sources of data were provided in the database. Statistical and qualitative information were presented as a table (PDF files) and as a MS Access database. All files were posted on the NDL web site in March 2003. An example of the finished database is given in Fig. 3.

3. Results and discussion

Literature review: Literature searches produced approximately 475 articles published since 1970. There were about 275 articles on fruits, vegetables and beverages other than citrus fruits and tea and about 200 articles on citrus fruits and tea. A quick review of the articles revealed that about 125 articles contained analytical data. Other types of articles included general review (16), articles on antioxidant activity (14), and bioavailability (5) of the flavonoids, articles on analytical methods for identification only (80), and articles which reported values for other compounds, e.g., phenolic acids or total by the subclass of flavonoids (36). In addition to those articles more articles were excluded for miscellaneous reasons such as insufficient documentation, values reported on dry weight basis only, and articles in a language other than English (40). Approximately 158 articles out of 200 on citrus fruits and tea were eliminated for various reasons mentioned above. Further evaluation of 125 articles with analytical data resulted in the rejection of data from 28 articles for various reasons including unacceptable analytical methods and data for some food items not consumed by, or available to the general population. The database contains acceptable data extracted from 97 articles (Table 1).

Example of databae page

NDB NO.	Description	Subclass	Flavonoid	Mean	SE	N	Min	Max	CC	Sources of data
09050	Blueberries, raw	Anthocyanidins	Cyanidin	15.02	0.84	12	4.79	28.72	B	29
			Delphinidin	29.54	0.14	12	20.82	47.37	B	29
			Malvidin	49.21	1.42	12	32.95	69.44	B	29
			Peonidin	7.05	0.53	12	1.01	19.37	B	29
			Petunidin	11.73	0.09	12	7.19	18.25	B	29
		Flavan-3-ols	(-)-Epicatechin	1.11	0.00	4	1.11	1.11	B	5
			(-)-Epicatechin 3-gallate	0.00	0.00	4	0.00	0.00	B	5
			(-)-Epigallocatechin	0.00	0.00	4	0.00	0.00	B	5
			(-)-Epigallocatechin 3-gallate	0.00	0.00	4	0.00	0.00	B	5
			(+)-Catechin	0.00	0.00	4	0.00	0.00	B	5
			(+)-Gallocatechin	0.00	0.00	4	0.00	0.00	B	5
		Flavonols	Kaempferol	0.00	0.00	6	0.00	0.00	B	12, 33
			Myricetin	0.82	0.15	6	0.00	2.60	B	12, 33
			Quercetin	3.11	0.04	7	1.70	7.30	B	12, 33, 44

Fig. 3. Example of a database page.

This Special Interest Database contained flavonoid values for 225 foods including fruits, vegetables, herbs, and beverages. Most of the data in this database came from European sources. For example, most of the data on catechins came from the Netherlands (Arts et al., 2000a) whereas data on flavonols came from the Netherlands (Hertog et al., 1992), Finland (Häkkinen et al., 1999) and England (Price et al., 1998a, b) and thus, they may not necessarily be representative of flavonoid levels in the similar US foods. Quercetin, a flavonol, was the most ubiquitous compound in all foods. All of the five subclasses of compounds were not present in all the foods. Taxonomically related plants synthesize similar patterns of flavonoids with regard to specific components and relative concentrations (Robards and Antolovich, 1997). Typically, one or two subclasses predominated in a particular food. For example, flavonols are present in vegetables and herbs, flavanones in citrus fruits, anthocyanidins in colored berries, flavan-3-ols in tea and fruits like apples, cherries and apricots.

A survey of flavonoid reports indicated that authors did not analyze all the subclasses due to a lack of suitable specific analytical methods and standards. Recently published references by Merken and Beecher (2000) and Merken et al. (2001) report new methods that can analyze compounds from all the subclasses simultaneously using HPLC.

Data quality: Out of a total of 1469 observations on aggregated data, 3% (41) received the highest confidence code of A, 61% received B, while 31% were assigned C and 5% were given D. The flavonoids data for tea, particularly black tea, had the highest ratings due to achieving high ratings in all five evaluation categories. Dutch (Arts et al., 2000a, b) and Finnish (Häkkinen et al., 1999) data for selected foods received good ratings for both analytical method and analytical quality control and as a result had confidence codes of B or greater. In general, data with a confidence code C had low ratings for the sampling plan category and had ratings of "0" for the analytical quality control (QC). Typically D quality data did not contain sufficient information on sampling plan, analytical method or analytical QC. In the previous version of the data evaluation system, low scores for analytical method led to the exclusion of data from further compilation. However, in the flavonoid database we retained the data with low analytical method score so that all data could continue to be reviewed. Since this is a preliminary database, the authors expect the quality of individual estimates to improve as additional representative values become available. Also, data from some non-English language articles will be included in the first update of the database.

When the general quality of the data was analyzed for the different subclasses of the flavonoids, the subclass flavan-3-ols received the best ratings (5% A, 82% B, and 13%C and no D ratings). The ratings for the subclasses flavonols, flavones and flavanones were comparable with most of the data having the confidence codes of B or C in equal proportions. There were limited data for anthocyanidins (only 173 records for all 6 compounds), but the general quality of those data was good (41% B and 59% C) (Table 3). The possible reasons for the limited amount of data for anthocyanidins include: (1) anthocyanidins are not as widely distributed (mostly in the berries), (2) data generated by the pH differential method were excluded, and (3) data reported as "total anthocyanidins" were set aside.

Further analysis for the five evaluation categories by each subclass is illustrated in Table 4. Each category can achieve a maximum of 20 points, 0–5 considered as below average, 6–10 as average, 11–15 as above average and 16–20 as best. The ratings for the *sampling plan* were average varying from 8 to 10 for flavan-3-ols (9), flavonols (9), flavones (8) and flavanones (10). The

Table 3
Confidence Code (CC) ratings for the flavonoids by subclasses

Subclass	Confidence Code (CC)			
	A (%)	B (%)	C (%)	D (%)
Flavonols	1	45	46	8
Flavones	1	49	38	12
Flavanones	3	38	59	—
Flavan-3-ols	5	82	13	—
Anthocyanidins	—	41	59	—

Table 4
Ratings for evaluation categories by flavonoid subclasses

Subclass	Sampling plan average (min., max.)	Sample handling average (min., max.)	Analytical method average (min., max.)	Analytical QC average (min., max.)	Sample number average (min., max.)
Flavonols	9 (2, 14)	17 (12, 20)	11 (2, 15)	6 (0, 17)	14 (1, 20)
Flavones	8 (2, 14)	17 (12, 20)	10 (3, 15)	5 (0, 17)	9 (1, 20)
Flavanones	10 (2, 14)	17 (12, 20)	8 (2, 15)	3 (0, 17)	18 (1, 20)
Flavan-3-ols	9 (2, 14)	19 (12, 20)	12 (2, 15)	9 (0, 17)	16 (1, 20)
Anthocyanidins	12 (2, 14)	16 (14, 20)	6 (4, 13)	0 (0, 0)	17 (1, 20)

average rating for the anthocyanidins was 12. As mentioned earlier, ''the number of regions'' criterion for this category was modified, by substituting ''the number of countries'' for ''the number of regions'' in the aggregated data to accommodate the international nature of this database. If the number of regions (countries in this database) was $\geqslant 4$ the criterion, number of regions, received the highest rating of 10. While the total number of values for anthocyanidins were few (173), the sources of data were diverse, e.g., data for red wines came from Spain, Germany, France and the United States, and for various berries, the data came from Canada and the United States, thus contributing to the maximum score (10) for the number of regions criterion.

The *sample handling* category for all the subclasses received good ratings ranging from 12 to 20 with the average of 17. Most of the articles did not give information on moisture content of the samples. This information helps to assess the effect of the storage conditions for the samples and to provide a comparable basis for aggregating individual values. Similarly, very few studies reported the validation of homogenization procedures by analyzing different aliquots of the same homogenate. However, the large amount of data on tea, wines and juices resulted in higher average ratings for this category. Generally these foods do not need homogenization nor are the data on moisture content as critical to component values.

Ratings for the *analytical method* category for flavan-3-ols, flavonols and flavones were between 10 and 12 (above average), while they were average for flavanones (8) and for anthocyanidins (6). One reason for these ratings can be attributed to the lack of certified or standard reference materials for flavonoids. In addition, the absence of details about variability of the analytical process (%CVs) and evidence of recovery studies for component levels affected the ratings for this category.

For the category, *analytical quality control*, only the data for the flavan-3-ols subclass had an acceptable rating (9.5) due to the large amount of data on vegetables, fruits and beverages from the Netherlands (Arts et al., 2000a, b) and the data on teas from the United States (Unilever Bestfoods, North America, unpublished data). Both sources reported use of an in-house quality control material with good results for day-to-day precision (CV 5%–10%). For other subclasses the ratings varied between 3 and 6. It is important for analysts to use an in-house quality control material made specifically for the study and sample matrix to ensure comparability of day-to-day precision and repeatability of the analyses. As mentioned previously, assessment of accuracy is

Table 5
Foods with appreciable amounts of flavonoids (Single data source)

Food description	Compounds with appreciable amounts (mg/100 g)
Apples, raw, with skin	Catechin (0.95), Epicatechin (8.14), Quercetin (4.42)
Apricots, raw	Catechin (14.95), Epicatechin (6.06), Quercetin (2.55)
Blackberries, raw	Epicatechin (18.08)
Blueberries, raw	Cyanidin (15.02), Delphinidin (29.54), Malvidin (49.21), Peonidin (7.05), Petunidin (11.73)
Cherries, sweet, raw	Cyanidin (111.43), Peonidin (5.15), Quercetin (1.25)
Cranberries, raw	Epicatechin (4.20), Quercetin (14.02), Anthocyanidins[a]
Grapefruit, raw	Hesperetin (1.5), Naringenin (53.00)
Grapes, black, raw	Catechin (8.94), Epicatechin (8.04), Anthocyanidins[a]
Raspberries, raw	Cyanidin (42.17), Delphinidin (0.50), Malvidin (1.23), Pelargonidin (3.70), Catechin (0.97), Epicatechin (8.26),
Broad beans, immature seeds, raw	Catechin (12.83), Epicatechin (22.51), Epigallocatechin (14.03), Myricetin (2.60), Quercetin (2.00)
Onions, red, raw	Cyanidin (13.14)
Peppers, all hot varieties, raw	Luteolin (1.34-6.93), Quercetin (0.51-50.63)
Spices, dill weed, fresh	Isorhamnetin (43.50), Kaempferol (13.33), Myricetin (0.70), Quercetin (55.15)
Spices, thyme, fresh	Apigenin (0.50), Luteolin (51.00)
Buckwheat flour, whole groat	Epicatechin (3.53), Quercetin (2.72)
Chocolate bar, dark	Catechin (11.90), Epicatechin (41.50)
Chocolate bar, milk	Catechin (2.90), Epicatechin (10.45)

[a]Anthocyanidins are expected, but no values were available.

limited without CRMs or SRMs. This category is important, particularly in the absence of the certified or standard reference materials.

The range of ratings averaged by subclass for the *number of samples* category was good (9–18). However, values for some food items came from a single source. For example, anthocyanidin values for blueberries ($n = 12$), and cherries ($n = 7$) and catechin values for apples ($n = 28$) came from single studies. Twelve different cultivars of blueberries (Gao and Mazza, 1994) and 7 different cultivars of cherries (Gao and Mazza, 1995) were analyzed in respective studies. Seven different varieties of apples from four different seasons were analyzed for catechins by Arts et al. (2000b) thus making the number of samples analyzed 28. These compounds are prominent in these foods and more data from other sources (laboratories) could confirm the values and then improve the confidence codes. Data for onions (yellow), tea and red wine were abundant and found to be of good quality. Quercetin, the most prominent flavonol in onions received a confidence code of B based on 294 analytical samples from nine different sources. Also, there were ample data for tea (black and green) and red wine. The data received confidence codes of B. Most of the data for other fruits and vegetables came from single sources. Table 5 shows some commonly consumed fruits, vegetables, and other foods that contain appreciable amounts of flavonoids for which values came from single sources. It would be desirable to have more data for these foods from various countries.

Preliminary review of available data indicated considerable variation in the flavonoid content in some foods. Flavonoid compounds are often produced by plants in response to stress. Plant

diseases, insects, climate, ultraviolet radiation, etc. may cause stress (Robards and Antolovich, 1997). Other sources of variability include cultivars, growing location, agricultural practices, processing and storage conditions. Observations on some raw data showed seasonal (e.g., black grapes and broad beans, Arts et al., 2000b) and cultivar (e.g., cherries, Gao and Mazza (1995) and blueberries, Gao and Mazza, 1994) differences. In some cases, mean values for individual flavonoids in a particular food came from different data sources that were compiled to generate a mean value. However, most of the values were based on a limited number of samples. Furthermore, users of the data should exercise caution when comparing flavonoid values for different forms of a food, such as between raw and cooked forms of the same food. As with any nutrient database, values for different forms of the food may have been collected from different sources. If a value in the cooked food is less than in the raw food, it does not necessarily mean that the particular flavonoid was reduced by cooking. This kind of comparison is valid only when the paired samples are prepared and analyzed in the same laboratory.

4. Conclusions

This is a preliminary Special Interest Database on monomeric compounds in the five subclasses of dietary flavonoids: flavonols, flavones, flavanones, flavan-3-ols and anthocyanidins. It provides the first compilation of 26 flavonoids of interest in these five subclasses in 225 foods. The overall quality of the data was good with 64% of the observations receiving B or better confidence codes. The quality of data for the subclass flavan-3-ols was better than others (5% A, 82% B and 13% C). Most of the data came from the European countries, the Netherlands, Finland, Spain, UK, and some from the United States and Canada. Therefore the data may not be representative of the flavonoid content of US foods. More data are needed on the foods mentioned in Table 5, particularly for anthocyanidins in different colored berries. The database provides a relative indication of flavonoid patterns to be expected in different food classes. It also provides the preliminary and limited statistics on possible variability in levels of flavonoids and the basis for setting future priorities for research. The USDA's ARS has recently completed a study of flavonoid levels in 59 fruits, vegetables and nuts procured from nationwide sampling to generate representative values for US foods. These data will be compared to those compiled from the earlier literature sources. Finally, specific points on evaluation of data quality may help future analysts to generate high quality data for flavonoids in foods.

References

Arts, I.C.W., van de Putte, B., Hollman, P.C.H., 2000a. Catechin content of foods commonly consumed in the Netherlands. 2. Tea, wine, fruit juices, and chocolate milk. Journal of Agricultural and Food Chemistry 48, 1752–1757.

Arts, I.C.W., van de Putte, B., Hollman, P.C.H., 2000b. Catechin content of foods consumed in the Netherlands. 1. Fruits, vegetables, staple foods, and processed foods. Journal of Agricultural and Food Chemistry 48, 1746–1751.

Formica, J.V., Regelson, W., 1995. Review of the biology of quercetin and related bioflavonoids. Food and Chemical Toxicology 33, 1061–1080.

Gao, L., Mazza, G., 1995. Characterization, quantitation, and distribution of anthocyanins and colorless phenolics in sweet cherries. Journal of Agricultural and Food Chemistry 43, 343–346.

Gao, L., Mazza, G., 1994. Quantitation and distribution of simple and acylated anthocyanins and other phenolics in blueberries. Journal of Food Science 59, 1057–1059.

Häkkinen, S.H., Kärenlampi, S.O., Heinonen, I.M., Mykkänen, H.M., Törrönen, A.R., 1999. Content of flavonols quercetin, myricetin, and kaempferol in edible berries. Journal of Agricultural and Food Chemistry 47, 2274–2279.

Heinonen, I.M., Meyer, A.S., Frankel, E.N., 1998. Antioxidant activity of berry phenolics on human low-density lipoprotein and liposome oxidation. Journal of Agricultural and Food Chemistry 46, 4107–4112.

Hertog, M.G.L., 1996. Epidemiological evidence on potential health properties of flavonoids. Proceeding of the Nutrition Society 55, 385–397.

Hertog, M.G.L., Hollman, P.C.H., Katan, M.B., Kromhout, D., 1993a. Intake of potentially anticarcinogenic flavonoids and their determinants in adults in the Netherlands. Nutrition and Cancer 20, 21–29.

Hertog, M.G.L., Hollman, P.C.H., Katan, M.B., 1993b. Content of potentially anticarcinogenic flavonoids in tea infusions, wines, and fruit juices. Journal of Agricultural and Food Chemistry 41, 1242–1246.

Hertog, M.G.L., Hollman, P.C.H., Katan, M.B., 1992. Content of potentially anticarcinogenic flavonoids of 28 vegetables and fruits commonly consumed in the Netherlands. Journal of Agricultural and Food Chemistry 40, 2379–2383.

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., Beecher, G.R., 1987. A system for evaluating the quality of published nutrient data: selenium, a test case. Food Nutrition Bulletin 9 (Suppl.), 177–193.

Hollman, P.C.H., Bijsman, M.N.C.P., van Gameren, Y., Cnossen, E.P.J., de Vries, J.H.M., Katan, M.B., 1999. The sugar moiety is a major determinant of the absorption of dietary flavonoid glycosides in man. Free Radical Research 9 (Suppl.), 569–573.

Justesen, U., Knuthsen, P., Leth, T., 1997. Determination of plant polyphenols in Danish foodstuffs by HPLC-UV and LC-MS detection. Cancer Letters 114, 165–167.

Kühnau, J., 1976. The flavonoids: a class of semi-essential food components: their role in human nutrition. World Review of Nutrition 24, 117–120.

Le Marchand, L., Murphy, S.P., Hankin, J.H., Wilken, L.R., Kolonel, L.N., 2000. Intake of flavonoids and lung cancer. Journal of National Cancer Institute 92, 154–160.

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 American Dietetic Association 93, 284–296.

Merken, H.M., Merken, C.D., Beecher, G.R., 2001. Kinetics method for the quantitation of anthocyanidins, flavonols, and flavones in foods. Journal of Agricultural and Food Chemistry 49, 2727–2732.

Merken, H.M., Beecher, G.R., 2000. Liquid chromatographic method for the quantification of prominent flavonoid aglycones. Journal of Chromatography A 897, 177–184.

Middleton Jr., E., 1996. Biological properties of plant flavonoids: an overview. International Journal of Pharmacognosy 34, 344–348.

Price, K.R., Casuscelli, F., Colquhoun, I.J., Rhodes, M.J.C., 1998a. Composition and content of flavonol glycosides in broccoli florets (*Brassica oleracea*) and their fate during cooking. Journal of the Science of Food Agriculture 77, 468–472.

Price, K.R., Colquhoun, I.J., Barnes, K.A., Rhodes, M.J.C., 1998b. Composition and content of flavonol glycosides in green beans and their fate during processing. Journal of Agricultural and Food Chemistry 46, 4898–4903.

Rauha, J.-P., Remes, S., Heinonen, M., Hopia, A., Kähkönen, M., Kujala, T., Pihlaja, K., Vuorela, H., Vourela, P., 2000. Antimicrobial effects of Finnish plant extracts containing flavonoids and other phenolic compounds. International Journal of Food Microbiology 56, 3–12.

Rice-Evans, C., Spencer, J.P.E., Schroeter, H., Rechner, A., 2000. Bioavailability of flavonoids and potential bioactive forms in vivo. Drug Metabolism and Drug Interaction 17, 291–310.

Rice-Evans, C., Miller, N.J., Bolwell, P.G., Bramley, P.M., Pridham, J.B., 1995. The relative antioxidant activities of plant-derived polyphenolic flavonoids. Free Radical Research 22, 375–383.

Robards, K., Antolovich, M., 1997. Analytical chemistry of fruit bioflavonoids. A review. Analyst 122, 11R–34R.

Unilever Bestfoods, North America. Unpublished data for tea.