

Process of formulating USDA's Expanded Flavonoid Database for the Assessment of Dietary intakes: a new tool for epidemiological research

Seema A. Bhagwat^{1*}, David B. Haytowitz¹, Shirley I. Wasswa-Kintu² and Pamela R. Pehrsson¹

¹Nutrient Data Laboratory, ARS, USDA, 10300 Baltimore Avenue, Building 005, Beltsville, MD 20705, USA

²Department of Nutrition and Food Science, University of Maryland, College Park, MD 20742, USA

(Submitted 17 June 2014 – Final revision received 4 February 2015 – Accepted 10 April 2015 – First published online 29 June 2015)

Abstract

The scientific community continues to be interested in potential links between flavonoid intakes and beneficial health effects associated with certain chronic diseases such as CVD, some cancers and type 2 diabetes. Three separate flavonoid databases (Flavonoids, Isoflavones and Proanthocyanidins) developed by the USDA Agricultural Research Service since 1999 with frequent updates have been used to estimate dietary flavonoid intakes, and investigate their health effects. However, each of these databases contains only a limited number of foods. The USDA has constructed a new Expanded Flavonoids Database for approximately 2900 commonly consumed foods, using analytical values from their existing flavonoid databases (Flavonoid Release 3.1 and Isoflavone Release 2.0) as the foundation to calculate values for all the twenty-nine flavonoid compounds included in these two databases. Thus, the new database provides full flavonoid profiles for twenty-nine predominant dietary flavonoid compounds for every food in the database. Original analytical values in Flavonoid Release 3.1 and Isoflavone Release 2.0 for corresponding foods were retained in the newly constructed database. Proanthocyanidins are not included in the expanded database. The process of formulating the new database includes various calculation techniques. This article describes the process of populating values for the twenty-nine flavonoid compounds for every food in the dataset, along with challenges encountered and resolutions suggested. The new expanded flavonoid database released on the Nutrient Data Laboratory's website would provide uniformity in estimations of flavonoid content in foods and will be a valuable tool for epidemiological studies to assess dietary intakes.

Key words: Flavonoids; Database expansion; Calculation techniques

Dietary flavonoids are secondary plant metabolites, and are present in fruits, vegetables, spices, herbs, nuts, legumes and their products. Recently, several systematic reviews with meta-analyses of cohort and/or case–control studies have supported the expected beneficial effects of flavonoids on chronic diseases like CVD⁽¹⁾, type 2 diabetes⁽²⁾ and some cancers^(3,4); the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort study, conducted in several centres across Europe, has observed reduced risk of colorectal cancer⁽⁵⁾, oesophageal cancer⁽⁶⁾ and gastric adenocarcinoma⁽⁷⁾. Recently, two other studies^(8,9) observed a lower risk of type 2 diabetes associated with higher consumption of anthocyanins. Although one study⁽¹⁰⁾ was not able to associate total flavonoid intakes with reduced risk of stroke, it did observe that increased intakes of flavanones, a subclass of

flavonoids, were inversely associated with reduced risk of ischaemic stroke. Quercetin, a flavonol, was found to be associated with reduced risk of gastric cancer in a case–control study⁽¹¹⁾. Thus, flavonoids have displayed diverse bioactivities associated with different compounds.

Food composition databases for dietary flavonoids have been found to be necessary tools to support these research studies. The Nutrient Data Laboratory (NDL) of the Agricultural Research Service at the US Department of Agriculture (USDA), developed a database for isoflavones, 'USDA Database for the Isoflavone Content of Selected Foods', in 1999⁽¹²⁾ and updated it in 2008, Release 2, (IDB 2.0)⁽¹³⁾. In 2003, the NDL developed a database for flavonoids, 'USDA Database for the Flavonoid Content of Selected Foods' for predominant dietary flavonoids in five subclasses (flavonols,

Abbreviations: EPIC, European Prospective Investigation into Cancer and Nutrition; FDB 3.1, USDA Database for the Flavonoid Content of Selected Foods, Release 3.1; FDB-EXP, USDA's Expanded Flavonoid Database for the Assessment of Dietary Intakes; FG, Food Group; FNDDS 4.1, Food and Nutrient Database for Dietary Studies, Release 4.1; IDB 2.0, USDA Database for the Isoflavone Content of Selected Foods, Release 2.0; NDB, Nutrient Data Bank; NDL, Nutrient Data Laboratory; SR, standard reference; SR22, National Nutrient Database for Standard Reference, Release 22; USDA, US Department of Agriculture.

* **Corresponding author:** S. A. Bhagwat, fax +1 301 504 0632, email seema.bhagwat@ars.usda.gov

flavones, flavanones, flavan-3-ols and anthocyanidins)⁽¹⁴⁾; it was updated in 2007⁽¹⁵⁾, 2011⁽¹⁶⁾, and most recently in 2013, Release 3.1 (FDB 3.1)⁽¹⁷⁾. Together, these two databases encompass all of the predominant dietary flavonoids except proanthocyanidins. The NDL is in the process of updating the USDA database for proanthocyanidins⁽¹⁸⁾. Flavonoid values in all these databases are reported as mg/100 g edible portion on fresh weight basis. In 2013, the NDL formulated a new Flavonoids Database for approximately 2900 commonly consumed foods, ‘USDA’s Expanded Flavonoid Database for the Assessment of Dietary Intakes’ (FDB-EXP)⁽¹⁹⁾, using analytical values from FDB 3.1 and IDB 2.0 as the foundation to calculate values for all the twenty-nine flavonoid compounds included in these two databases. Original analytical values in Flavonoid Release 3.1 and Isoflavone Release 2.0 for corresponding foods were retained in the newly constructed database. Values were calculated only for the foods and/or the flavonoid compounds that were found missing in these two databases. Unlike FDB 3.1 and IDB 2.0 which do not contain analytical values for all the twenty-six compounds and all the three isoflavones respectively for every food, FDB-EXP contains values for all the twenty-nine flavonoid compounds for every food, which are a combination of analytical and calculated values. It does not currently include proanthocyanidin values. NDL plans to add proanthocyanidin data to the expanded database after it is updated. The database comprises a subset of foods included in NDL’s National Nutrient Database for Standard Reference 22 (SR22)⁽²⁰⁾ released in 2009. These 2926 foods represent a snapshot of foods commonly consumed in the United States that was used to develop Food and Nutrient Database for Dietary Studies 4.1 (FNDDS 4.1)⁽²¹⁾, and was used to analyse data from What We Eat in America component of the National Health and Nutrition Examination Survey 2007–8⁽²²⁾. Table 1 provides the number of foods, subclasses of flavonoids, and compounds in each subclass included in FDB 3.1, IDB 2.0 and the newly constructed FDB-EXP databases. The FDB-EXP is available on the NDL’s website: <http://www.ars.usda.gov/nutrientdata/ flav>.

As of 2013, our ongoing literature searches revealed that these flavonoid databases, singly or together, have been used for more than seventy studies conducted in the USA, Europe, the UK and Australia to estimate flavonoid intakes, and to evaluate associations between flavonoid intakes and various chronic diseases. Although the databases were developed after comprehensive literature searches for analytical data, and the analysis of nationally representative samples of fruits, vegetables and nuts in the USA at the USDA’s laboratories^(23,24), the databases do not include every food (fruit, vegetable, herb or legume) expected to contain some flavonoid compounds. In addition, these databases, with the exception of the IDB 2.0 which contains some prepared foods, do not include multi-ingredient prepared foods/food products, although these foods may contain ingredients that contribute flavonoids to those dishes. Therefore, scientists who use these databases for epidemiological studies need to calculate flavonoid contents for food items unavailable in these databases and for the multi-ingredient foods, to avoid underestimation of intakes.

This article describes the process of constructing a new Flavonoids Database for approximately 2900 commonly consumed foods. Analytical values from FDB 3.1 and IDB 2.0 are used as the foundation to calculate values for all the twenty-nine flavonoid compounds included in these two databases. Also described are the challenges encountered during this process, and how the NDL addressed these challenges.

Materials and methods

The general approach to developing USDA’s flavonoids databases is described in the earlier publication⁽²⁵⁾. FDB 3.1 contains data for twenty-six flavonoid compounds in five subclasses of flavonoids: flavonols, flavan-3-ols, flavones, flavanones and anthocyanidins in 306 foods, although every food may not have values for all the twenty-six selected flavonoids. IDB 2.0 contains data for three prominent isoflavones: daidzein, genistein and glycitein, in 549 foods, but may not have values for all the three compounds for all the foods. These two databases

Table 1. US Department of Agriculture databases for flavonoid subclasses as foundation for the newly constructed flavonoids database

Database	Year	Number of foods	Class of compounds	Names of compounds
Isoflavones (IDB 2.0) ^{*(13)}	2008	549	Isoflavones	Daidzein, genistein and glycitein
Flavonoids (FDB 3.1) ^{*(17)}	2013	506	Flavonols	Isorhamnetin, kaempferol, myricetin and quercetin
			Flavones	Apigenin and luteolin
			Flavanones	Eriodictyol, hesperetin and naringenin
			Flavan-3-ols	Catechin, epicatechin, epicatechin-3-gallate, epigallocatechin, epigallocatechin-3-gallate, galocatechin, theaflavin, theaflavin-3-gallate, theaflavin-3'-gallate, theaflavin-3,3'-digallate and thearubigins
			Anthocyanidins	Cyanidin, delphinidin, malvidin, pelargonidin, peonidin and petunidin
Expanded flavonoids (FDB-EXP) ^{*(19)}	2014	2926	All the six classes listed in this column	All the twenty-nine compounds listed in this column

IDB 2.0, USDA Database for the Isoflavone Content of Selected Foods, Release 2.0; FDB 3.1, USDA Database for the Flavonoid Content of Selected Foods, Release 3.1; FDB-EXP; USDA’s Expanded Flavonoid Database for the Assessment of Dietary Intakes.

* <http://www.ars.usda.gov/nutrientdata/ flav>

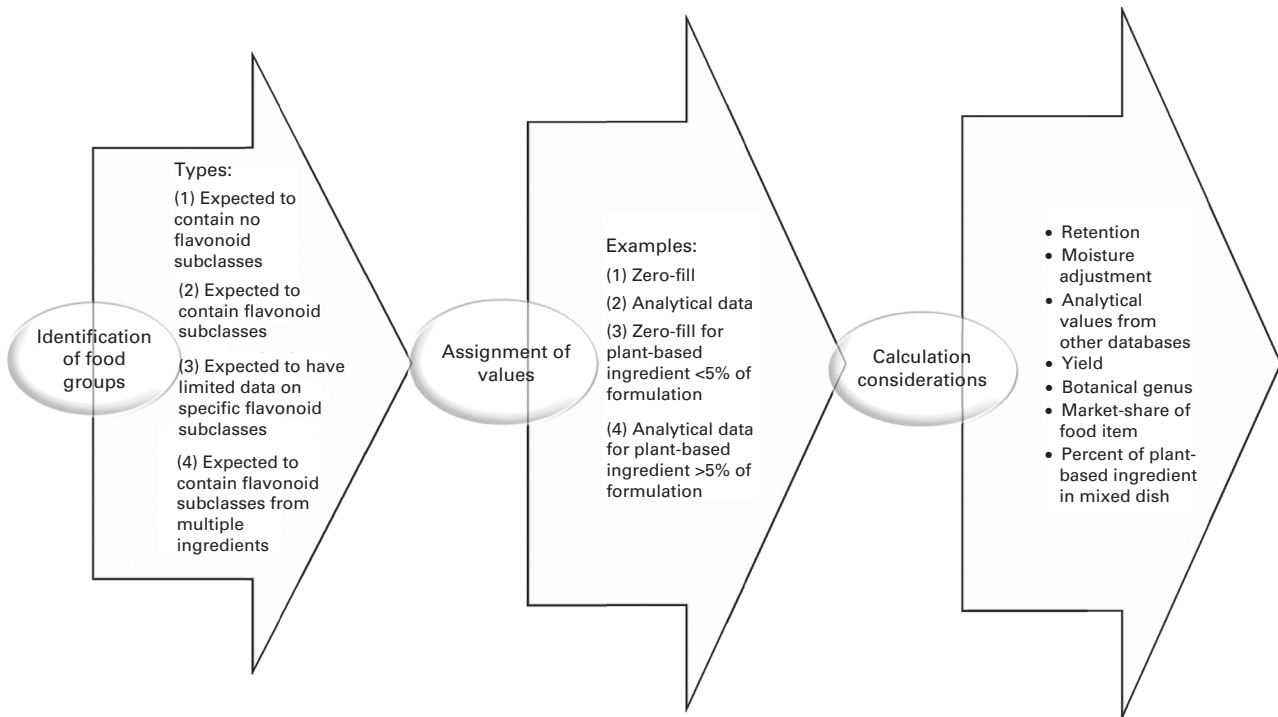


Fig. 1. Process flow for populating flavonoid profiles.

contain data for twenty-nine predominant dietary flavonoids in six subclasses. The NDL has observed that scientists generally analyse only the expected predominant class of compounds in a particular food type; for example, flavanones in citrus fruits or anthocyanidins in coloured berries. The objective of the expansion project was to populate full flavonoid profiles for the twenty-nine flavonoid compounds for each food in a subset of 2926 foods in NDL's SR22 that was used to develop FNDDS 4.1, using analytical values from FDB 3.1 and IDB 2.0 as foundations. The process involved various steps including (1) assigning logical zero values, (2) matching analytical values, (3) calculating flavonoid values (moisture adjustment, using retention factors for cooked/processed foods, food yield factors and substituting values with similar foods) and (4) calculating values for multi-ingredient foods (Fig. 1). Each step is described in detail as follows.

Assigning logical 'zero' values

Standard reference (SR) contains twenty-five Food Groups (FG) covering a wide variety of foods consumed in the United States. Animals cannot synthesise flavonoids like plants but get them through foods that they consume⁽²⁶⁾. It may be possible that animals can accumulate flavonoids in their tissues through their feed. Based on the limited data in FDB 3.1 and IDB 2.0, and an extensive review of literature, flavonoids data for animal product food groups were not found except for a very small amount of isoflavones in eggs⁽²⁷⁾, and for 'total isoflavones' in some dairy, seafood and meat products⁽²⁸⁾. The USDA databases accept values reported for individual compounds only, not if reported only as 'total'. Therefore, a '0' value was assigned to all the

flavonoids for foods in animal product groups (beef, pork, poultry, finfish, shellfish, dairy, fats, sausages and luncheon meats), and multi-ingredient foods derived from these food groups, if other ingredients were not expected to contain flavonoids. Vegetable oils were also assumed to have no flavonoids since no data were available except for olive oils, which are extracted from fruits and which did have analytical values for flavones. More comprehensive data on these foods are needed to fully characterise their flavonoids content.

Only one or two subclasses of flavonoids are predominant in most food items: for e.g. flavanones are a major subclass in citrus fruits and may contain some flavonols. Therefore, 'zero' values were assigned to flavonoid compounds in the subclasses that were not expected to be present in a particular food or FG. Table 2 illustrates the general scheme of assuming logical zeroes in each FG.

Matching analytical values

FG which are expected to contain one or more flavonoid subclasses included fruits/fruit juices, vegetables/vegetable products, spices/herbs, nuts/seeds and legumes. During the development of FDB 3.1 and IDB 2.0, foods were assigned Nutrient Data Bank numbers (NDB), unique five-digit numerical codes used in SR, if the food descriptions matched with those in SR. If exact matches were not available, temporary NDB numbers were assigned. The analytical flavonoid values for the exact matches for these NDB numbers in FDB 3.1 and IDB 2.0 with NDB numbers in FDB-EXP were retained. Food descriptions for the temporary NDB numbers were also reviewed to find matches for similar foods in the FDB-EXP.

Table 2. Assigning logical zero values

Food group	Number of foods	Food group description	Flavonols	Flavones	Flavanones	Flavan-3-ols	Anthocyanidins	Exceptions
0100	188	Dairy and egg products	zx	z	z	zx	z	Products with added chocolate and fruit
0200	36	Spices and herbs					zx	Wine vinegar
0300	272	Baby foods						
0400	117	Fats and oils	z	z	z	z	z	
0500	101	Poultry products	z	z	z	z	z	
0600	148	Soups, sauces and gravies	zx	zx	z	z	z	Vegetable soups and sauces
0700	87	Sausages and luncheon meats	z	z	z	z	z	
0800	198	Breakfast cereals		z	z		zx	Added fruits and nuts
0900	197	Fruits and fruit juices						
1000	120	Pork products	z	z	z	z	z	
1100	302	Vegetables and vegetable products			zx			Naringenin in tomatoes
1200	47	Nut and seed products		z	z		zx	Cyanidin in some nuts
1300	92	Beef products	z	z	z	z	z	
1400	136	Beverages						
1500	94	Finfish and shellfish products	z	z	z	z	z	
1600	101	Legumes and legume products		z	z			
1700	65	Lamb, veal and game products	z	z	z	z	z	
1800	230	Baked products	zx	zx	zx		zx	Added fruits, nuts, chocolate and soya flour
1900	175	Sweets			zx			
2000	69	Cereal grains and pasta		z	z		z	
2100	47	Fast foods						≥ 5% flavonoid containing ingredient
2200	18	Meals, entrees and side dishes						≥ 5% flavonoid containing ingredient
2500	84	Snacks						≥ 5% flavonoid containing ingredient
1950	2	Gums	z	z	z	z	z	

z denotes assumed zero; zx denotes mostly assumed zero with some exception.

Calculating flavonoid values

After assigning available analytical values to those foods which were direct matches by food descriptions, various techniques were then employed to calculate values for the remaining compounds and/or foods according to the procedures described by Schakel *et al.*⁽²⁹⁾. The various procedures used to calculate values where analytical values were unavailable are described below.

Moisture adjustment. For plant-based foods that undergo moisture changes due to cooking, drying or dilution, values for full flavonoid profiles were calculated from another form of the same food or from a similar food. A moisture factor based on the change in total solids was applied (e.g. raw to cooked asparagus or fresh to dried basil). If the moisture content was not reported in the published article, moisture values from the SR were used for the same food or sometimes for similar food.

Retention factors. To account for changes in flavonoid contents after cooking/processing of the foods, it was necessary to determine retention factors. Literature searches done while developing the flavonoid databases had retrieved only fifteen studies that analysed vegetables in raw and cooked

forms^(30–44). Most of them analysed only flavonols, one analysed apigenin⁽³⁵⁾ and one naringenin in Brussels sprouts⁽⁴¹⁾. Different cooking methods such as boiling, steaming, frying or microwaving were used in these studies. The average retention for flavonols, flavones and flavanones from these studies was approximately 86% (58–132%). The cooking methods were not taken into consideration when the average retention percentage was calculated. There were no reports on anthocyanidins retention. After consulting scientists at the Food Composition and Methods Development Laboratory of Agricultural Research Service/USDA, as well as data from limited literature sources, retention factors of 85% for flavonols, flavanols, flavanones and flavones and 50% for anthocyanidins (because of their heat labile characteristics) were established for the FDB-EXP. The application of dry heat (i.e. baking or air drying) was considered to have negligible effects on flavonoid losses (J Harnly, personal communication, 2012). Therefore, retention factors were not applied when drying process was used, e.g. for fresh to dried fruits or fresh to dried herbs. If the analytical values were available for cooked/processed foods, these values were retained in preference to those calculated by applying retention factors. Retention factors were not needed for isoflavones, as analytical

values were available for most of the raw and cooked/processed foods that contained isoflavones. True retention factors can be generated only by paired studies by analysing half of the food sample raw and the other half after cooking, and recording weights of the raw and cooked samples. Cooked or canned plant-based foods without full flavonoid profiles were calculated from values for raw forms of the same food, using estimated retention factors to account for the loss of flavonoids during processing.

The following example illustrates how the moisture and retention factors were applied to calculate unavailable quercetin values for cooked beets from the analytical values for raw beets.

Formula: $N_t = (N_s \times S_s/S_t) \times R$, where N_t = the nutrient (quercetin) content of the target item (cooked beets); N_s = the nutrient (quercetin) content of the source item (raw beets = 0.134 mg/100 g); S_s = the total solid content (total weight – moisture content) of the source item (raw beets = 100 – 87.58 = 12.42 g/100 g); S_t = the total solid content (total weight – moisture content) of the target item (cooked beets = 100 – 87.06 = 12.94 g/100 g); R = retention factor (estimated retention factor for quercetin 0.85).

Using the above formula, quercetin content of cooked beets is calculated thus:

$$N_t = (0.134 \times 12.42)/(12.94) \times 0.85 = 0.12 \text{ mg}/100 \text{ g.}$$

Food yield factors. Food yield factors were also applied to account for food processing effects in instances where values were available only for a different form of the same food. For example, for canned foods, food yield factors applied to the raw form to adjust for yields of solid foods after draining liquids ranging from 53–67% (USDA, 1975)⁽⁴⁵⁾, since the flavonoids were expected only in the fruit or vegetable (a yield factor of 67% was applied to canned peaches, i.e. a can of peaches contained 67% fruit and 33% liquid). Retention factors were applied to compensate for processing losses. Although some flavonoids are leached in the canning liquid (water or syrup)⁽⁴⁶⁾, no adjustments were made for these losses.

Substitution with similar foods. Values from a similar food were substituted for the food or compound considering botanical origins, or other similarities like colour and texture. For example, flavan-3-ols, flavanones, anthocyanidins and flavonol myricetin values of fresh blackberries were used for fresh mulberries, without any adjustments for the lack of analytical values for these compounds in mulberries.

Market share. Other factors such as market share data were considered when estimating flavonoid values for some items. The market shares were considered only for wines and grapes because of the lack of specific NDB numbers distinguishing red and white wines, or red and green grapes. The NDB no. 14084 in SR describes a table wine without specifying whether it is red or white. However, this item is not included in the FDB 3.1, as data were available for both red and white wines, and as flavonoids values, particularly that of anthocyanidin, would be quite different. Market share proportions of 47% red wine, 40% white wine and 13% blush wine were reported by AC Nielsen supermarket

data, 2009 and the US Department of Commerce, 2011⁽⁴⁷⁾. However, due to lack of analytical data for flavonoid for blush wines (only one source), a generic profile for table wine (NDB no. 14084) was created by using 50% of red and 50% of white wine values. In the development of FDB 3.1, some food items (e.g. red and green grapes) were assigned more specific provisional NDB numbers; it was done to differentiate them from each other on the basis of flavonoid content, although grapes have a single NDB no. (09132) in SR that does not distinguish red and green grapes. Due to lack of nationwide market share or consumption data on table grapes specified as red or green separately, a generic value for all grapes, corresponding to the NDB no. 09132 in SR, was also developed by using both the red (50%) and green (50%) grape values.

Generic profiles. A generic profile for twenty-nine flavonoid compounds was prepared for common leafy vegetables, using values for nine leafy vegetables from FDB 3.1. This profile was used to estimate values for a specific leafy vegetable, when flavonoid values were not available for any or some compounds. A generic profile for fruits was also developed using fifteen fruits, and used for less common fruits when similar fruits were not available for substitution of flavonoid values or missing values.

Other sources. Values for instant tea powders provided in this database were drawn from the unpublished figures provided by the Unilever Lipton Company in 2002. The weight for added flavouring and/or artificial sweetener like saccharin was disregarded when calculating flavonoid values for tea powders with flavours and/or with sugar substitutes, i.e. the same values were used for unsweetened/unflavoured and sweetened with saccharin/flavoured instant tea powders and prepared teas as well.

Values from other databases such as Phenol-Explorer, Release 2⁽⁴⁸⁾ were used for a limited number of foods/compounds not available in the FDB 3.1. For instance, kaempferol values for most nuts were obtained from Phenol-Explorer.

Calculating flavonoid values for multi-ingredient foods

Multi-ingredient foods with one or more ingredients of plant origin such as baby foods, soups, breakfast cereals, beverages and prepared meals/entrees may contain some flavonoids, depending on amounts of plant-based ingredients. However, few multi-ingredient food items were included in FDB 3.1 because of the lack of analytical data. For multi-ingredient foods, formulations developed by NDL scientists⁽⁴⁹⁾ were used to estimate percentages of flavonoid-containing ingredients. These formulations use regression equations which take the ingredient lists on the labels of these products, plus the nutrient content of the ingredients from SR, and estimate the proportion of each ingredient in the food item. The formulations were developed by food group specialists in NDL who are familiar with the products; but they were developed for other nutrients, and occasionally were not suitable for flavonoid calculations and hence had to be modified suitably. In general, the flavonoid-containing ingredient was used in calculations only if it contributed $\geq 5\%$ of the total by



Table 3. Examples of calculation techniques and factors used

Calculation technique	Food with no values	Source of food with values	Flavonoid compound	Factor used
Moisture factor	Dried basil	Fresh basil	All	11.29
Retention factor	Collards, cooked, boiled and drained	Collard, raw	Flavonols	0.85
Yield factor	Pears, canned and drained	Pears, raw	All	0.67
Substitution with similar food	Mulberries, raw	Blackberries, raw	Flavan-3-ols, flavanones, anthocyanidins, and myricetin	1.00
Formulation for mixed dish	Pizza with meat and vegetables	Sweet green peppers (5%)	All	0.05
		Tomato sauce (23%)		0.23
		Table olives (5%)		0.05
		Raw onions (5%)		0.05

weight. The exceptions were cocoa powder (regular and alkalisied) and soya protein isolates/soya flour because of their high contents of catechins and isoflavones respectively per unit weight. Multi-ingredient foods containing less than five percent of each plant-based ingredient were estimated to have no flavonoids present in them. Whenever orange juice was one of the ingredients ($\geq 5\%$), e.g. as in baby food juices or citrus juice drinks, we used values of the entry 'orange juice, chilled, includes from concentrate (NDB no. 09209)' to calculate flavanone values for these foods for consistency in estimation.

Table 3 illustrates some of the techniques used to calculate flavonoid values.

Results and discussion

The FDB-EXP database was released on the NDL's website in September 2014. It is available as Microsoft® ACCESS files, not as PDF files, which include descriptions of foods, food groups, flavonoids data, compound (nutrient) numbers, source codes and derivation codes. The 2926 foods in the FDB-EXP were distributed across twenty-five FG, and contained a total of 84854 data points. All the flavonoid values from FDB 3.1 and IDB 2.0 for corresponding foods were retained in FDB-EXP. Values were calculated only for foods and/or compounds that were not available in these two databases. The values in FDB-EXP are reported as mg/100 g edible portion. Since only five plant-based FG and foods that contain plant foods as ingredients are expected to have compounds from two or three flavonoids subclasses, and are likely to contain only some flavonoid compounds from each of those subclasses, each food item may have analytical values for only three to five flavonoid compounds. The five food groups expected to have some flavonoids comprise approximately 780 foods out of 2926 total foods in the FDB-EXP. Therefore, 73% of the data points (61943 out of a total of 84854 data points) received logical 'zero' values. Twenty-four percent values (20365) in the database were calculated by techniques described in the earlier methods section, and three percent (2546) were matched with the analytical values that were drawn from FDB 3.1 and IDB 2.0. In SR, a single food item can have many different forms, and each form will have a unique NDB number, e.g. a specific fruit can be fresh, frozen, dried, canned in water, canned in syrup or a vegetable

could be boiled without draining, with draining, with added salt and without added salt. Therefore, the number of data points in these five flavonoid-containing FG is very large. As analytical data may exist for only one of the forms, usually fresh (raw), it was necessary to calculate values for other forms of the foods. Therefore, it is reasonable to say that only a small percentage of data in the FDB-EXP are analytical.

The significant approach for expanding FDB 3.1 and IDB 2.0 flavonoid databases involved calculating values for unavailable analytical data after assigning logical or assumed 'zero' and matching analytical values. Using moisture factors, retention factors and yield factors required many algorithms.

The EPIC studies^(50,51) used FDB Release 2.1 (2007) and Phenol-Explorer, Release 2⁽⁴⁸⁾ to estimate intakes of flavanols, flavones, flavanones and anthocyanidins using retention factors of 70, 35 and 25% for fried, microwaved and boiled foods respectively for all flavonoid compounds. These factors were based on a single study⁽³²⁾. These dissimilar approaches which were applied to retention factors based on cooking methods by EPIC studies *v.* the same based on classes of flavonoid compounds by USDA will affect the values for cooked foods in these two databases, and consequently the estimation of flavonoid intakes.

Calculating flavonoid values for multi-ingredient foods like meals/entrees was another challenge. Although formulations developed by the NDL scientists⁽⁴⁹⁾ were applied to estimate percentages of flavonoid-containing ingredients for these foods, it required some more algorithms. Soya flour and soya protein isolates are used in the bakery products as dough conditioners⁽⁵²⁾. It was also confirmed by personal communication with manufacturers in the USA. Therefore, it is not surprising that the authors of the EPIC-Norfolk cohort study for the UK⁽⁵²⁾ observed that bread and bread rolls contributed the highest percentages of isoflavones in the diet of both men and women. Analytical values for total isoflavones in doughnut varied from 0.60 to 5.31 mg/100 g in IDB 2.0, depending on the kind of doughnut. Small isoflavones values were also reported in IDB 2.0 for eggs⁽²⁷⁾. The poultry industry uses soyabean meal as the main protein in poultry feed, and it comprises 66% of all the proteins used⁽⁵³⁾. Use of soyabean meal in poultry feed was also confirmed by personal communication (R Angel, Department of Poultry Science, University of Maryland, USA, 2013).

The unexpected isoflavone values in eggs may be due to the use of soyabean meal in the chicken feed⁽⁵⁴⁾.

Most foods, except for some fresh fruits and vegetables, are consumed after processing (different cooking methods, canning, drying, etc.), but no retention factors were available at the time the database was developed, to calculate values for the flavonoid compounds post-processing. Therefore, the major difficulty encountered during the expansion of the databases was the estimation of retention factors. The NDL in collaboration with Food Composition and Methods Development Laboratory has planned a study to generate true retention factors for different compounds/classes of flavonoids, using different cooking methods. Eggs are a staple food used in many recipes and are thus highly consumed. Analyses of eggs, some meat products and multi-ingredient foods would enhance these databases.

Summary

Plant foods such as fruits, vegetables, legumes and a few grains are the main sources of flavonoids in the human diet. Flavonoids data were not available for animal foods except for isoflavone values for eggs. Only two or three subclasses of flavonoids are predominant in most foods, e.g. flavanones in citrus. Generally, researchers focus on analysing the expected predominant compounds only in the food they are analysing. Therefore, analytical values for the full profile of all the twenty-nine flavonoid compounds included in the FDB-EXP were not available for many foods. In order to properly estimate flavonoid intakes, it is necessary to create full profiles of flavonoid compounds for each food item. The FDB-EXP contains data for twenty-nine flavonoids in six classes for approximately 2900 food items from SR22 which provides the basis for the FNDDS 4.1.

The FDB-EXP will be a valuable tool for epidemiologists to estimate flavonoid intakes of a population of interest, and study associations between intakes and various chronic diseases. The database contains flavonoid values for a large number of foods in many different forms that are commonly consumed in the United States. It can also be employed as an effective tool to address directly discrepancies created by using different techniques by different scientists while calculating data for missing compounds or foods in their own databases. Studies using FDB-EXP will have uniformity in the estimations of flavonoids contents in foods. As more and more analytical data for flavonoids become available through published literature the need to impute flavonoid values will diminish.

Limitations

The database for 84854 flavonoid values was formulated on the basis of analytical values directly available for just three percent of the data. Many assumptions were made in order to calculate those values for the rest of the data, depending on the experience of the authors in the database development, and consultations with experts in the areas of flavonoids analyses. The database contains twenty-nine selected

prominent flavonoid compounds for 2926 selected foods, but not other flavonoid compounds that could possibly be present in those foods.

Acknowledgements

The present study was supported by the Office of Dietary Supplements (ODS), NIH, Agreement no. Y1CN5010. The salary of S. I. W.-K. was supported by the University of Maryland, College Park, Maryland (Agreement no. 58-1235-1-98). The ODS and the University of Maryland had no role in the design, analysis or writing of the present article.

The authors' contributions are as follows: S. A. B., D. B. H. and S. I. W.-K. designed, carried out the study, and developed the Expanded Flavonoid Database; D. B. H. developed and executed the algorithms to calculate flavonoid values; S. I. W.-K. sketched the figure; S. A. B. wrote the manuscript; D. B. H., S. I. W.-K. and P. R. P. contributed significant comments. All authors read and approved the final manuscript.

None of the authors has any conflict of interest.

References

1. Wang X, Ouyang YY, Liu J, *et al.* (2014) Flavonoid intake and risk of CVD: a systematic review and meta-analysis of prospective cohort studies. *Br J Nutr* **111**, 1–11.
2. Babu PVA, Liu D & Gilbert ER (2013) Recent advances in understanding the anti-diabetic actions of dietary flavonoids. *J Nutr Biochem* **24**, 1777–1789.
3. Woo HD & Kim J (2013) Dietary flavonoid intake and smoking-related cancer risk: a meta-analysis. *PLOS ONE* **8**, e75604.
4. Hui C, Qi X, Qianyong Z, *et al.* (2013) Flavonoids, flavonoid subclasses and breast cancer risk: a meta-analysis of epidemiologic studies. *PLOS ONE* **8**, e54318.
5. Zamora-Ros R, Not C, Guino E, *et al.* (2013) Association between habitual dietary flavonoid and lignan intake and colorectal cancer in a Spanish case–control study (the Bellvitge Colorectal Cancer Study). *Cancer Causes Control* **24**, 549–557.
6. Vermeulen E, Zamora-Ros R, Duell EJ, *et al.* (2013) Dietary flavonoid intake and esophageal cancer risk in the European prospective investigation into cancer and nutrition cohort. *Am J Epidemiol* **178**, 570–581.
7. Zamora-Ros R, Agudo A, Lujan-Barroso L, *et al.* (2012) Dietary flavonoid and lignan intake and gastric adenocarcinoma risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Am J Clin Nutr* **96**, 1398–1408.
8. Jennings A, Welch AA, Spector T, *et al.* (2012) Intakes of anthocyanins and flavones are associated with biomarkers of insulin resistance and inflammation in women. *J Nutr* **144**, 202–208.
9. Wedick NM, Pan A, Cassidy A, *et al.* (2012) Dietary flavonoid intakes and risk of type 2 diabetes in US men and women. *Am J Clin Nutr* **95**, 925–933.
10. Cassidy A, Rimm EB, O'Reilly EJ, *et al.* (2012) Dietary flavonoids and risk of stroke in women. *Stroke* **43**, 946–951.
11. Ekstrom AM, Serafini M, Nyren O, *et al.* (2011) Dietary quercetin intake and risk of gastric cancer: results from a population-based study in Sweden. *Ann Oncol* **22**, 438–443.
12. U.S. Department of Agriculture, Agricultural Research Service (1999) USDA Database for the Isoflavone Content

- of Selected Foods. Updated version (release 2): <http://www.ars.usda.gov/nutrientdata/flav>
13. Bhagwat S, Haytowitz DB & Holden JM (2008) USDA Database for the Isoflavone Content of Selected Foods, Release 2.0. U.S. Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory Home Page: <http://www.ars.usda.gov/nutrientdata/flav> (accessed January 2014).
 14. U.S. Department of Agriculture, Agricultural Research Service (2003) USDA Database for the Flavonoid Content of Selected Foods. Updated version (release 3.1): <http://www.ars.usda.gov/nutrientdata/flav>
 15. U.S. Department of Agriculture, Agricultural Research Service (2007) USDA Database for the Flavonoid Content of Selected Foods, Release 2. Updated version (release 3.1): <http://www.ars.usda.gov/nutrientdata/flav>
 16. Bhagwat S, Haytowitz DB & Holden JM (Ret.) (2011) USDA Database for the Flavonoid Content of Selected Foods, Release 3. U.S. Department of Agriculture, Agricultural Research Service.
 17. Bhagwat S, Haytowitz DB & Holden JM (Ret.) (2013) USDA Database for the Flavonoid Content of Selected Foods, Release 3.1. U.S. Department of Agriculture, Agricultural Research Service. Nutrient Data Laboratory Home Page: <http://www.ars.usda.gov/nutrientdata/flav> (accessed January 2014).
 18. USDA Database for the Proanthocyanidin Content of Selected Foods. U.S. Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory Home Page: <http://www.ars.usda.gov/nutrientdata/flav> (accessed September 2012).
 19. Bhagwat S, Haytowitz DB & Wasswa-Kintu S (2014) USDA's Expanded Flavonoid Database for the Assessment of Dietary Intakes. U.S. Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory Home Page: <http://www.ars.usda.gov/nutrientdata/flav> (accessed September 2014).
 20. U.S. Department of Agriculture, Agricultural Research Service (2009) USDA National Nutrient Database for Standard Reference, Release 22. Nutrient Data Laboratory Home Page: <http://www.ars.usda.gov/nutrientdata> (accessed January 2014).
 21. U.S. Department of Agriculture, Agricultural Research Service (2010) USDA Food and Nutrient Database for Dietary Studies (FNDDS 4.1). Food Surveys Research Group Home Page: http://www.ars.usda.gov/main/site_main.htm?modecode=12-35-50-00 (accessed September 2012).
 22. Centers for Disease Control and Prevention, National Center for Health Statistics (2010) National Health and Nutrition and Examination Survey 2007–08 (NHANES 2007–2008). http://www.cdc.gov/nchs/nhanes/search/nhanes07_08.aspx (accessed September 2012).
 23. Harnly JM, Doherty RF, Beecher GR, *et al.* (2006) Flavonoid content of U.S. fruits, vegetables, and nuts. *J Agric Food Chem* **54**, 9966–9977.
 24. Gu L, Kelm MA, Hammerstone JF, *et al.* (2004) Concentrations of proanthocyanidins in common foods and estimations of normal consumption. *J Nutr* **134**, 613–617.
 25. Holden JM, Bhagwat SA, Haytowitz DB, *et al.* (2005) Development of a database of critically evaluated flavonoids data: application of USDA's data quality evaluation system. *J Food Comp Anal* **18**, 829–844.
 26. Peterson J & Dwyer J (1998) Flavonoids: dietary occurrence and biochemical activity. *Nutr Res* **18**, 1995–2018.
 27. Horn-Ross PL, Barnes S, Lee M, *et al.* (2000) Assessing phytoestrogen exposure in epidemiologic studies: development of a database (United States). *Cancer Causes Control* **11**, 289–298.
 28. Kuhnle GGC, Dell'Aquila C, Aspinall SM, *et al.* (2008) Phytoestrogen content of foods of animal origin: dairy products, eggs, meat, fish, and seafood. *J Agric Food Chem* **56**, 10099–10104.
 29. Schakel SF, Buzzard IM & Gebhardt S (1997) Procedures for estimating nutrient values for food composition databases. *J Food Comp Anal* **10**, 102–114.
 30. Andlauer W, Stumpf C & Furst P (2000) Influence of the acetification process on phenolic compounds. *J Agric Food Chem* **48**, 3533–3536.
 31. Chun OK, Smith N, Sakagawa A, *et al.* (2004) Antioxidant properties of raw and processed cabbages. *Int J Food Sci Nutr* **55**, 191–199.
 32. Crozier A, Lean MEJ, McDonald MS, *et al.* (1997) Quantitative analysis of the flavonoid content of commercial tomatoes, onions, lettuce, and celery. *J Agric Food Chem* **45**, 590–595.
 33. Ewald C, Fjellkner-Modig S, Johansson K, *et al.* (1999) Effect of processing on major flavonoids in processed onions, green beans, and peas. *Food Chem* **64**, 231–235.
 34. Fanasca S, Roupheal Y, Venneria E, *et al.* (2009) Antioxidant properties of raw and cooked spears of green asparagus cultivars. *Int J Food Sci Nutr* **44**, 1017–1023.
 35. Ferracane R, Pellegrini N, Visconti A, *et al.* (2008) Effects of different cooking methods on antioxidant profile, antioxidant capacity, and physical characteristics of artichoke. *J Agric Food Chem* **56**, 8601–8608.
 36. Hakkinen SH, Karenlampi SO, Mykkanen HM, *et al.* (2000) Influence of domestic processing and storage on flavonol contents in berries. *J Agric Food Chem* **48**, 2960–2965.
 37. Lombard K, Peffley E, Geoffriau E, *et al.* (2005) Quercetin in onion (*Allium cepa* L.) after heat-treatment simulating home preparation. *J Food Comp Anal* **18**, 571–581.
 38. Makris DP & Rossiter JT (2001) Domestic processing of onion bulbs (*Allium cepa*) and asparagus spears (*Asparagus officinalis*): effect on flavonol content and antioxidant status. *J Agric Food Chem* **49**, 3216–3222.
 39. Miglio C, Chiavaro E, Visconti A, *et al.* (2008) Effects of different cooking methods on nutritional and physicochemical characteristics of selected vegetables. *J Agric Food Chem* **56**, 139–147.
 40. Mullen W, Stewart AJ, Lean ME, *et al.* (2002) Effect of freezing and storage on the phenolics, ellagitannins, flavonoids, and antioxidant capacity of red raspberries. *J Agric Food Chem* **50**, 5197–5201.
 41. Pellegrini N, Chiavaro E, Gardana C, *et al.* (2010) Effect of different cooking methods on color, phytochemical concentration, and antioxidant capacity of raw and frozen *Brassica* vegetables. *J Agric Food Chem* **58**, 4310–4321.
 42. Price KR, Bacon JR & Rhodes MJC (1997) Effect of storage and domestic processing on the content and composition of flavonol glucosides in onion (*Allium cepa*). *J Agric Food Chem* **45**, 938–942.
 43. Price KR, Casascelli F, Colquhoun IJ, *et al.* (1998) Composition and content of flavonol glycosides in broccoli florets (*Brassica oleracea*) and their fate during cooking. *J Sci Food Agric* **77**, 468–472.
 44. Price KR, Colquhoun IJ, Barnes KA, *et al.* (1998) Composition and content of flavonol glycosides in green beans and their fate during processing. *J Agric Food Chem* **46**, 4898–4903.
 45. U.S. Department of Agriculture, Agricultural Research Service (1975) Food yields, summarized by different stages of preparation. In *Agricultural Handbook No. 102*. Washington, DC: US Department of Agriculture.

- <http://www.ars.usda.gov/SP2UserFiles/Place/80400525/Data/Classics/ah102.pdf>
46. Rickman JC, Barrett DM & Bruhn CM (2007) Nutritional comparison of fresh, frozen and canned fruits and vegetables. Part 1. Vitamins C and B and phenolic compounds. *J Sci Food Agric* **87**, 930–944.
 47. Hodgen DA (2011) *U.S. Wine Industry – 2011*. Washington, DC: U.S. Department of Commerce.
 48. Neveu V, Perez-Jimenez J, Vos F, *et al.* (2010) Phenol-Explorer: an online comprehensive database on polyphenol contents in foods. *Database (Oxford)* **2010**, bap024.
 49. Haytowitz DB, Lemar LE & Pehrsson PR (2009) USDA's Nutrient Databank System – a tool for handling data from diverse sources. *J Food Comp Anal* **22**, 433–441.
 50. Zamora-Ros R, Knaze V, Lujan-Barroso L, *et al.* (2011) Estimated dietary intakes of flavonols, flavanones and flavones in the European Prospective Investigation into Cancer and Nutrition (EPIC) 24 hour dietary recall cohort. *Br J Nutr* **106**, 1915–1925.
 51. Zamora-Ros R, Knaze V, Lujan-Barroso L, *et al.* (2011) Estimation of the intake of anthocyanidins and their food sources in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Br J Nutr* **106**, 1090–1099.
 52. Mulligan AA, Kuhnle GG, Lentjes MA, *et al.* (2013) Intakes and sources of isoflavones, lignans, enterolignans, coumestrol and soya-containing foods in the Norfolk arm of the European Prospective Investigation into Cancer and Nutrition (EPIC-Norfolk), from 7 d food diaries, using a newly updated database. *Public Health Nutr* **16**, 1454–1462.
 53. Waldroup PW & Smith K (ND) *Fact sheet: Soybean Use – Poultry*. Ankeny, IA: Soybean Meal Information Center.
 54. Saitoh S, Sato T, Harada H, *et al.* (2001) Transfer of soy isoflavone into the egg yolk of chickens. *Biosci Biotechnol Biochem* **65**, 2220–2225.