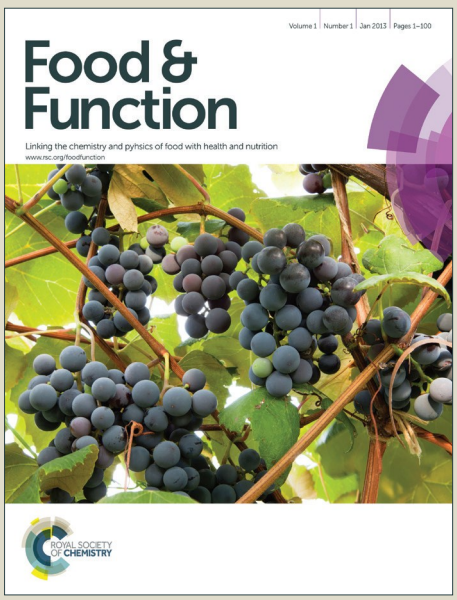
Food & Function

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/foodfunction

Comparison of flavonoid intake assessment methods.

Kerry L Ivey, Kevin Croft, Richard L Prince, Jonathan M Hodgson

Author affiliations: Harvard T. H. Chan School of Public Health, Department of Nutrition, Boston, USA [KLI]; Harvard T. H. Chan School of Public Health, Department of Epidemiology, Boston, USA [KLI]; South Australian Health and Medical Research Institute, Infection and Immunity Theme, School of Medicine, Flinders University, Adelaide, Australia [KLI]; University of Western Australia, School of Medicine and Pharmacology, Sir Charles Gairdner Hospital Unit, Nedlands, Australia [RLP]; Sir Charles Gairdner Hospital, Department of Endocrinology and Diabetes, Nedlands, Australia [RLP]; University of Western Australia, School of Medicine and Pharmacology, Royal Perth Hospital, Perth, Australia [JMH KC].

Address for correspondence: Kerry L Ivey: Harvard T. H. Chan School of Public Health, Department of Nutrition, Building 2, Level 3, 677 Huntington ave, Boson, Massachusetts, USA, 02155; Email: kivey@hsph.harvard.edu.

Short title: Flavonoid intake assessment.

ABSTRACT

Background: Flavonoids are a diverse group of polyphenolic compounds, found in high concentrations in many plant foods and beverages. High flavonoid intake has been associated with reduced risk of chronic disease. To date, population based studies have used the United States Department of Agriculture (USDA) food content database to determine habitual flavonoid intake. More recently, a new flavonoid food content database, Phenol-Explorer (PE), has been developed. However, the level of agreement between the two databases has yet to be explored.

Aim: To compare the methods used to create each database, and to explore the level of agreement between the flavonoid intake estimates derived from USDA and PE data.

Design: The study population included 1 063 randomly selected women aged over 75 y. Two separate intake estimates were determined using food composition data from the USDA and the PE databases.

Results: There were many similarities in methods used to create each database, however, there are several methodological differences that manifested in differences in flavonoid intake estimates between the 2 databases.

Despite differences in net estimates, there was a strong level of agreement between total-flavonoid, flavanol, flavanone and anthocyanidin intake estimates derived from each database. Intake estimates for flavanol monomers showed greater agreement than flavanol polymers. The level of agreement between the two databases was weakest for the flavonol and flavone intake estimates.

Conclusion: In this population, application of USDA and PE source data yielded highly correlated intake estimates for total-flavonoids, flavanols, flavanones and anthocyanidins. For these sub-classes, the USDA and PE databases may be used interchangeably in

epidemiological investigations. There was poorer correlation between intake estimates for flavonols and flavones due to differences in USDA and PE methodologies. Individual flavonoid compound groups that comprise flavonoid sub-classes had varying levels of agreement. As such, when determining the appropriate database to calculate flavonoid intake variables, it is important to consider methodologies underpinning database creation, and which foods are important contributors to dietary intake in the population of interest.

BACKGROUND

Flavonoids represent a diverse group of polyphenolic compounds derived from a flavan (2-phenylchroman) nucleus (**Figure 1**). Derivations of this basic structure arise due to alterations in the 2(3) carbon-carbon bond, the formation of a ketone at carbon 4, and hydroxylation of carbons at various locations on the flavan backbone [1]. It is these derivations that give the over 4000 flavonoid molecules to be grouped into one of five main flavonoid sub-classes; flavonols, flavanols, flavones, flavanones, and anthocyainidins (**Table 1**)[2-4]. Each flavonoid sub-class is comprised of numerous individual compounds with varying degrees of polymerisation, glycosylation, hydroxylation and esterification.

Clinical trial and experimental data suggests a promising role of flavonoids in improving numerous chronic disease risk factors[5-11]. However, individual observational epidemiological studies of flavonoids and flavonoid rich foods have observed beneficial, null, or at times inconsistent or weaker than expected associations of flavonoid intake with risk of chronic disease outcomes [12-15]. In addition to natural variation on the flavonoid composition of foods, the validity of food composition databases used to compute flavonoid intake from traditional dietary assessment methodologies is likely to have an impact on estimated flavonoid intake, and henceforth observed associations, in the population[16].

It is extremely difficult to assess the validity of flavonoid intake assessment tools as there is currently no gold standard biomarker of total-flavonoid intake. As such, the use of food composition databases in conjunction with traditional dietary assessment methods provides the most practical means for analysis of habitual dietary flavonoid consumption in the community. In previous epidemiological studies, the most widely adopted flavonoid food content database has been that developed by the United States Department of Agriculture (USDA)[17 18]. Traditionally, food content data from the USDA database has been combined with food intake data from 24-hour recall and food frequency questionnaires (FFQ) to

estimate flavonoid intake [19 20]. It is these computed estimates of flavonoid intake which are used as the independent variable to investigate diet-disease relationships in population based settings.

More recently, the Phenol-Explorer[21] database has emerged, providing an additional high quality summary of the flavonoid content of commonly consumed foods. We, and others, have utilized this data in isolation, and in conjunction with USDA data, for assessing diet-disease relationships[22-24]. However, the manner in which the two databases deal with the complexity of flavonoid structure and content, and the degree to which these two databases agree when applied to a validated dietary assessment method has not yet been investigated. As such, this study aims to compare the methods used by each database to derive food content values, and to explore the level of agreement between the flavonoid intake estimates derived from USDA and PE data.

METHODS: METHODOLOGICAL COMPARISON

Due to the unique chemical properties of flavonoid compounds, methods used to create flavonoid food composition databases can dramatically affect the final estimates of the flavonoid content of food. The majority of naturally occurring flavonoids in foods are in the conjugated form, where variety of compounds, predominantly sugars, are bound to the flavan nucleus (**Figure 1**). These conjugates with sugars are called glycosides, and conjugate-free compounds are called aglycones.

There are distinct methodologies that can be utilised to determine the flavonoid concentration of food. One analytical approach is to use acidic or enzymatic hydrolysis to cleaves the conjugates from the flavonoid nucleus, to obtain aglycone compounds which are then subsequently quantified. This type of analysis 'summarises' all conjugates together in a single aglycone value. The USDA database primarily contains this type of values. The other analytical approach separates all the different conjugates in a food sample, and then quantifies each structurally unique compound individually. The data thus obtained are presented in the Phenol Explorer database.

What effect these methodological differences have on population level intake estimates is as yet unknown, and it is the purpose of this paper to quantify and describe differences in flavonoid intake estimates derived using food composition data from the USDA database ('USDA database for the flavonoid content of selected foods release 3.1' [25], 'USDA database for the proanthocyanidin content of selected foods'[18]) and from the PE database (Phenol-Explorer version 2.1[21]). Methodological data was obtained from associated documentation, and where information was missing or required clarification, the authors were contacted directly.

METHODS: FLAVONOID INTAKE COMPARISON

Participants

1,136 postmenopausal women above the age of 75 years were recruited into the Calcium Intake Fracture Outcome Age Related Extension Study in 2003. These participants had previously completed a 5-year prospective, randomized, controlled trial of oral calcium supplements to prevent osteoporotic fractures[26]. This study was approved by the Human Ethics Committee of the University of Western Australia, and written informed consent was obtained from all participants.

A total of 1,063 participants had complete food frequency and beverage intake data at baseline (2003). Participants had a mean age of 80 (±3) years, and a mean body mass index of 27 (±5) kg/m$^{2.}$

Dietary assessment

Baseline dietary intake was assessed using a validated semi-quantitative (FFQ) developed by the Anti-Cancer Council of Victoria [27-29]. Dietary intakes in g/day were estimated based on frequency of consumption and an overall estimate of usual portion size [30]. A beverage questionnaire was used to assess average tea and coffee consumption over the preceding 12 months.

Flavonoid intake

The method of computing flavonoid content of foods has been previously described in Ivey et al.[31], and is similar to the method adopted by many other investigators when assessing flavonoid intake [19].

Extraction procedures for the 2 databases (USDA and PE) were identical and were carried out by the same investigator. Both of these databases were used to derive two separate estimates of flavonoid intake: flavonoid intake based on food composition data from the USDA

(Flavonoid$_{USDA}$) and flavonoid intake based on food composition data from the PE (Flavonoid$_{PE}$) database.

The sum of assessed flavonoids for each flavonoid sub-class was calculated by summing the individual compounds of each flavonoid sub-class in the form expressed in each individual database. The terminology and classification systems used by each database varied; therefore, a standardized classification system was adopted throughout this study. The term 'anthocyanins' in this study refers to the PE anthocyanidin sub-class. The term flavanol in this study encapsulates both the Flavan-3-ol and Proanthocyanidin sub-classes in the USDA database. As such, flavanol$_{USDA}$ represented the sum of Proanthocyanidin polymer values in conjunction with total flavan-3-ol content. Proanthocyanidin monomer data from the USDA was not included in this analysis. We only utilised PE data that represented glycosides or aglycones in the form that they are naturally occurring in food. Specifically, in cases where HPLC data without hydrolysis was available, it was utilised.

The chalcone, dihydrochalcone and dihydroflavonol content of foods are described in the PE but not the USDA database. These compounds are typically considered a precursor to many flavonoid compounds, and not a flavonoid specifically. As such, these compounds were omitted from flavonoid intake computations. Isoflavones are commonly included in epidemiological analyses of total-flavonoid intake [32], and are included in the flavonoid section of the PE database. However, rather than sharing the nuclear structure of flavonoids (2-phenylchroman), isoflavones have a 3-Phenylchroman base structure. As such, isoflavones did not meet the criteria for inclusion in this comparison.

Flavonoid sub-class intake in mg/d was calculated by multiplying the estimated intake (g edible portion/d) from the FFQ and beverage questionnaires, with the flavonoid sub-class content (mg/g edible portion) of each food item on the questionnaires. Where multiple varieties of a food listed in the FFQ were reported in the databases, the average flavonoid

content of all similar varieties was computed, consistent with the descriptors used in the FFQ output. Foods in the FFQ that were not in the flavonoid databases entered as zero values for that particular database.

For those compounds with functional groups attached, the mass of the individual flavonoid conjugates have been incorporated into the food composition estimates in the Phenol-Explorer database. As such, we expressed each conjugate as the aglycone using molecular weight computations. The aglycone parent compounds were limited to only the parent compounds of the respective flavonoid sub-classes, and did not include aglycone masses for non-flavonoid compounds, such as gallic acid.

Statistics

Paired sample t-test was performed in order to compare the mean total-flavonoid intake estimates derived from the two databases. The extent to which estimates from USDA and PE databases are linearly related was explored with product-moment correlation coefficients. The relationship between measurement error and the mean estimated intake was gauged from the Bland and Altman plot. Spearman rank correlation coefficient was used to investigate the linear relationship between total-flavonoid intake groups based on tertiles of estimated intake from the two databases.

As a post hoc analysis, all investigations were repeated on the five flavonoid sub-classes, in order to explore the potential contribution of individual sub-classes to observed results. All data was analysed on SPSS (version 20; IBM, New York, NY) according to a pre-specified protocol.

RESULTS: METHODOLOGICAL COMPARRISON

Data sources

The PE database accessed only published data sources. The USDA included data from non-peer reviewed sources. Although these sources are cited, it is not apparent the extent by which industry contributed to the non-published food composition values. In terms of the analytical methods used by data sources, both databases were based on chromatography values. However, PE also included spectrophometric data.

Phenol-Explorer excluded studies from the database if it was deemed that an inappropriate method of polyphenol extraction was used or lack on information was provided on the method. They also excluded studies if there was a lack of information on phenolic standards used for quantification, mean content values without a description of the number of samples analysed, or content values reported in a graph. It is unclear if these factors were considered in USDA database development. Both databases only included values for specific flavonoid compounds, and omitted summary values for total flavonoid or flavonoid sub-class content of particular food items.

The PE database omitted data of non-edible parts of plants and non-commercial or experimental products. With the exception of low moisture content cereal products where standard moisture content data was used to convert results to a fresh weight, data sources that did not describe the moisture content of dried samples were excluded. A lack of descriptive data on the nature of the samples analysed was also a criteria for exclusion. USDA does list food sample criteria; however concurrence is inferred by methods of data display and aggregation.

Chemistry of included compounds (Supplementary Tables 1-5)

With the exception of (-)-epigallocatechin 3-gallate, theaflavin-3,3'-digallate, theaflavin-3-gallate and theaflavin-3'-gallate, the weight of flavonoid compounds in USDA food items was expressed as aglycone. Conversely, PE gives contents as individual glycosides and esters. In this analysis we first converted the Phenol Explorer compounds to the aglycone form, and the values of identical aglycones were subsequently summarised. Cis and trans isomers are not separately identified in the PE, and are instead represented as a total value. USDA does not list isomer criteria.

The authors identify that the 'USDA database for the proanthocyanidin content of selected foods'[18] is a provisional database, and has recognised that accuracy of food composition data in this area is limited by limitations in technical assessment methods.

Food content data

The flavonoid compound quantity of each food item is displayed as the mean weighted for the number of samples used to generate each original data. In calculating the mean content values, the PE also considered analytical methods used by the data sources by grouping the individual data points based on analytical technique prior to data aggregation. Food items are described using the USDA National Nutrient Database for Standard Reference in the USDA database, and the PE database used LanguaL descriptors. In both databases, the flavonoid content of solid foods is expressed as mg/100g of fresh weight of edible portion of food. Beverage data is expressed as mg/100mL in the PE, and as mg/100g adjusted by specific gravity in the USDA. The USDA standardised tea infusion data to 1% infusion strength, however did not adjust for brewing time. PE does not describe infusion strength considerations.

When data sources reported food items containing trace amounts of flavonoid compounds, the USDA estimated flavonoid content by multiplying the limit of quantitation (if available)

208 by 0.71. PE does not describe method for dealing with trace data. Zero values in both
209 databases represent true zeroes or levels below the limit of detection, whereas missing values
210 indicate an absence of available data.

RESULTS: FLAVONOID INTAKE COMPARRISON

Absolute flavonoid intake estimates (Table 2)

Mean total-flavonoid intake of the cohort estimated from USDA was 71% greater than that estimated from the PE. On a sub-class level, the USDA estimates were larger than the PE estimates for the flavanol, flavanone, and anthocyanidin sub-classes, and the PE estimates were greater than the USDA estimates were larger for the flavonol and flavone sub-classes.

The greatest proportional difference between the two databases was observed with the anthocyanidin sub-class, with USDA estimates being 8 times greater than the PE anthocyanidin intake estimates. When exploring anthocyanidin compound groups, the USDA and PE estimates were similar for pelargonidin compounds (4±4 and 3±5 mg/d, respectively). However, the mean estimates for cyaniding/peonidin and the delphinidin/malvidin/petunidin compounds were significantly greater ($P<0.05$) with the USDA data (59±52 and 55±57 mg/d, respectively) than with the PE data (3±4 and 5±8 mg/d, respectively). Furthermore, bananas contributed 4±3 mg/d to the anthocyandin$_{USDA}$ intake estimates, whereas bananas did not contribute to the anthocyandin$_{PE}$ intake estimate.

The greatest difference in absolute (mg/d) intake estimates was observed with the flavanol sub-class, with mean USDA estimates being 339 mg/d greater than that using the PE data. This difference is predominantly driven by the thearubigin compound group which accounted for 158±110 mg/d in the USDA estimates. The mean intake estimate for the theaflavin group of compounds was also greater when using the USDA as opposed to the PE data; 185±127 and 117±80 mg/d, respectively. As thearubigins compounds are not included in the PE database, thearubigins did not contribute to PE flavanol intake estimates. The mean intake estimates for the non-thearubigin polymeric flavanol compound group was similar for both USDA and PE; 184±146 and 179±122 mg/d, respectively.

235 It is important to note that the flavonoid-class intake estimates are not globally greater when
236 using the USDA data. The PE estimated daily flavonol intake was 350% (74 mg/d) greater
237 than the USDA estimated intake. At a compound level, this difference is largely explained by
238 the mean PE intake estimates of kaempferol and quercetin/isorhamnetin compound groups
239 (39±25 and 62±35 mg/d, respectively) being greater ($P<0.05$) than those derived using the
240 USDA database (8±4 and 22±11 mg/d, respectively). On the other hand, the mean myricetin
241 intake was lower 4 mg/d lower unsing the USDA data. Despite the USDA database omitting
242 information on 5 different groups of flavonols compounds (**Supplementary Table 1**), this
243 only contributed, on a mean level, < 1 mg/d to the difference between USDA and PE
244 flavonols intake estimates.

245 On an individual basis, there were no participants with identical intake estimates for
246 total-flavonoids from the USDA and PE databases. Similarly, all values for the estimates of
247 flavonoid sub-class intake were different between the 2 databases, with the exception of 2
248 participant estimates for flavone intake, and 1 participant estimate for flavanone intake.

249 From a population perspective, results of the paired sample t-test indicate that the mean
250 consumption estimates for total-flavonoid, as well as all flavonoid sub-classes, derived from
251 the USDA database were different to those derived from the PE database.

252 ***Extent to which the USDA and PE intake estimates are related***

253 The estimates of total-flavonoid intake from USDA and PE databases were strongly, linearly
254 and positively correlated (**Figure 2**). However, there was great heterogeneity in the level of
255 agreement for both the flavonoid sub-classes as well as the individual groups of compounds
256 contributing to each sub-class.

257 There was a near perfect linear association observed between the flavanone intake estimates
258 from the USDA and PE databases (**Figure 2e**). This strong agreement extended to all the

main flavanone compound groups assessed in both databases (**Supplementary Table 4**). Conversely, the strength of relation was weakest for the flavone intake estimates (**Figure 2d**) as well as its associated compound groups (**Supplementary Table 3**).

Although there was a strong linear relationship between the 2 databases for flavonol, flavanol, and anthocyanidin sub-class intake estimates, the agreement in estimates for the compound groups comprising these sub-classes were variable. One factor that may explain level of agreement is the number of datapoints contributing to an intake estimate. Compound groups with few FFQ food items with substantive concentrations in food, such as the flavanol compounds found mainly in tea (**Supplementary Table 3**), showed very strong agreement between the 2 databases.

Although impacting absolute intake estimates, the inclusion of thearubigins in the USDA database, but not in the PE database, did not substantively contribute to level of agreement in the intake estimates. The Pearson correlation coefficient for the USDA and PE intake estimates for polymeric flavonols was 0.75 ($P<0.001$). The level of agreement was not substantively improved by excluding thearubigins from the USDA intake estimate, where the resulting Pearson correlation coefficient for non-thearubigin flavanol polymers was 0.76 ($P<0.001$).

Relationship between level of disagreement and mean estimated intake **(Figure 3)**

The Bland and Altman plot demonstrates that the variability between the total-flavonoid estimates computed from the 2 databases is related to the size of the mean flavonoid intake estimates. The size of the difference between the two databases was greatest at higher levels of mean total-flavonoid intake. Similarly, the difference between estimates from the two databases was also proportional to the mean estimate for all flavonoid sub-classes.

Extent to which the USDA and PE intake estimates classify participants as low, moderate or high flavonoid consumers (**Table 3**)

In order to explore the ability of the two databases to rank intake of participants appropriately, we trichotomised the cohort into three levels of intake (low, moderate and high) based on tertiles of USDA and PE derived intake estimates. Despite substantial differences in participant classification, the classification of participants based on intake estimates derived from both databases were significantly correlated.

Ranking over 80% of participants identically, the reliability of the intake rankings from the two databases was high for total-flavonoid, flavanol and flavanone intake estimates. With less than 50% placed into identical intake groups, the USDA and PE databases showed poorest ranking agreement for the flavone sub-class.

DISCUSSION

This study aimed to compare the methods used by the USDA and PE databases to derive food content values. The majority of methodologies adopted by both databases were comparable. However, the major difference between the USDA and PE databases is the use of a different analytical approach to derive food content data. Values appearing in the USDA are obtained after hydrolysis of conjugates, whereas the PE contains separate data of individual conjugates. A drawback of hydrolysis might be that degradation or incomplete hydrolysis may occur that will lead to underestimation of the flavonoid content. This drawback is not present in the data of PE. However, various conjugates of a specific flavonoid may be present in a food, which each have to be quantified individually. Some of these conjugates may escape detection or quantification, because concentrations may be low, or proper standards are lacking. In the hydrolysis approach all the separate analytical signals are summarised into one bigger signal, the aglycone.

Neither database specifically identified the isomeric state of included compounds, primarily due to lack of source data for separate enantiomers. This omission is unlikely to affect the net estimates, as different isomeric forms are likely included in imputations of larger compound groups. However, absorption and bioactivity of flavonoids may be affected by isomeric form[33], as such the variability in bioactivity may not be reflected in current databases.

A notable difference between methodologies adopted by the two databases is in the data aggregation methods. In calculating the mean content values, the PE also considered analytical methods used by the data sources by grouping the individual data points based on analytical technique prior to data aggregation, whereas the USDA did not. The USDA imputed content data containing trace amounts of flavonoid compounds. It is unclear how the PE deals with trace data. The databases used different methods for expressing flavonoid concentration of beverages, and the USDA adopted a standardised infusion strength for tea,

318 whereas PE did not describe a standardisation method. Although potentially affecting content
319 estimates of individual flavonoid compounds in individual food items, this lack of
320 methodological congruency is unlikely to result in systematic differences between food
321 content estimates, and is unlikely to result in net alterations in flavonoid intake estimates.

322 By applying these food composition values to a validated FFQ, we then aimed to explore the
323 level of agreement between the flavonoid intake estimates derived from USDA and PE
324 databases. When applied to ACCV-FFQ food intake data, the mean total-flavonoid$_{USDA}$
325 intake estimate being 25% greater than that derived from PE data. Despite a high degree of
326 heteroscedasticity, the intake values derived from the two databases were related in a linear
327 fashion, in which 96% of the variation in intake estimates being explained by the relationship
328 between USDA and PE data. These differences in total-flavonoid intake estimates can be
329 explained by differences in flavonoid sub-class intake estimates, that arise as a result of
330 differences in methodologies adopted by each database to derive food content estimates,
331 flavonoid compounds assessed by each database, as well as the food items expressed in each
332 database.

333 The mean flavonol$_{PE}$ intake of the cohort was nearly 350% greater than the flavonol$_{USDA}$
334 estimate. This difference may be, in part, due to fact that the PE database provides data for
335 five additional groups of flavonol compounds which were not expressed in USDA.
336 Furthermore, the USDA database does not include the flavonol content data of chocolate. It is
337 these methodological differences that have contributed to the relatively poorer database
338 agreement when compared to other flavonoid sub-classes.

339 The major contributor to total-flavonoid intake was the flavanol sub-class; comprising 80%
340 of total-flavonoid$_{USDA}$ and 67% total-flavonoid$_{PE}$. With the mean estimated daily
341 flavanol$_{USDA}$ intake being more than 200% greater than the flavanol$_{PE}$ estimate, there was
342 substantial absolute difference in estimated flavanol intake. This is likely due to the absence

343 of thearubinigin data in the PE data. Data from other tea consuming nations show
344 thearubigins make a substantial contribution to total flavonoid intake[34], and in our study,
345 thearubigins made up 24% (158 mg/d) of total flavanol intake. When thearubigins are
346 excluded from analysis, mean $flavanol_{USDA}$ is 508 mg/d, compared to the 398 mg/d estimated
347 with PE data. Despite differences in absolute estimates, there was good correlation in
348 estimates from both databases, which likely arises due to the overrepresentation of tea in
349 flavanol intake estimates. The large contribution of tea-flavanols to total flavanol intake
350 means results in lower degree of variance attributable to differences between the two
351 databases in flavanol content estimates of flavanol containing foods. As flavanols contribute
352 substantially to total flavonoid intake, the high agreement between flavanol intake estimates
353 is reflected in the high agreement in total-flavonoid intake estimates.

354 Flavones are the best described sub-class in terms of food content estimates, likely because
355 this sub-class was included in the earliest flavonoid food content databases [35]. Despite this,
356 the $flavone_{PE}$ intake estimate is more than 3 times greater than the $flavone_{USDA}$ estimate, and
357 there was poor agreement between the two estimates, with a substantial degree of
358 heteroscedasticity. The higher PE food content estimates for food such as tea and fruit juice
359 likely contributed to the absolute differences and poor agreement between the two databases.
360 However, at a more global level, the differences in flavone intakes between USDA and PE
361 are probably caused by an analytical problem. A particular property of flavones is that they
362 contain C-glycosides in addition to the regular O-glycosides normally present. These C-
363 glycosides are hard to hydrolyse, and thus the data in the USDA database are expected to be
364 lower than PE as the PE database does not have this drawback as the food composition values
365 used are from analyses that quantified C-glycosides without hydrolysis.

366 Flavanone intake between the two databases is strongly correlated and similar in terms of net
367 estimates. The main reason for this is likely because both databases identify citrus and fruit

368 juice as being the major food source. Additionally, the flavanone content of other food items
369 is not well defined by each database, therefore, any variation attributable to differences in
370 food content estimates is minimal.

371 Although the two databases show moderately strong agreement in their ability to rank
372 participants as low, moderate and high anthocyanidin consumers, the daily USDA
373 anthocyanidin intake estimate was 800% (77 mg/d) greater than PE estimates. The difference
374 in net intake estimates is likely explained by the higher anthocyanidin food content values in
375 the USDA compared to the PE database. Through examination of data points contributing to
376 food content values, one can make inferences about the potential for measurement error. For
377 example, in this analysis, bananas made a small contribution to the higher anthocyanidin$_{USDA}$
378 estimates, and the validity of the USDA banana anthocyanidin estimates have previously
379 been raised[36].

380 The comparison of flavonoid intake estimates derived from each database is affected by the
381 foods included in the FFQ as well as the dietary intake pattern of the population investigated.
382 We have previously used the ACCV-FFQ to identify cross-sectional and prospective
383 flavonoid-disease relationships [29 31], which has been validated in populations of similar age
384 and geographical location to our cohort [37]. However, dietary patterns vary across different
385 geographical locations and age groups, and foods identified in FFQs typically reflect the
386 dietary pattern of the population to which it is administered. As such, the level of agreement
387 between the USDA and PE databases observed in this study may be different to the level of
388 agreement in cohorts of different nationality and age. With the flavonoid content of food
389 varying dramatically with factors pertaining to cultivars, growing conditions, food
390 processing, geography and season[38-40], the discrepancy between flavonoid intake estimates,
391 from either database, and 'true' flavonoid exposure, may also vary in different populations.

As there is no validated biomarker of long term, habitual, total flavonoid or flavonoid sub-class intake, it is not yet possible to determine which database provides intake estimates that most closely reflect reality. Having said this, by considering underlying methodologies and examining level of agreement between the 2 databases, one can indirectly infer the degree of measurement error associated with estimates derived from one, or both, of the databases. From the Bland and Altman plots, it is evident that for both total-flavonoids and flavonoid sub-classes, the level of disagreement between the databases increases with increasing intake estimates, suggesting that the potential for measurement error is greatest at higher absolute intakes. Furthermore, through examination of individual groups of flavonoid compounds, we observed that for some flavonoid-sub-classes, such as flavonols, flavanols, and anthocyanins, better agreement between the 2 databases can be achieved by looking at particular groups of flavonoid compounds making up that class. However, this is not the case for all sub-classes, and agreement was not substantively improved by exploring the compound-level associations.

The USDA and PE databases represent two comprehensive indexes of the flavonoid content of food. Due to differences in the methodologies underpinning their construction, the food composition data provided by each of the databases are not identical, and both provide unique glimpses into distinct aspects of the flavonoid content of food. At a population level, these methodological differences manifest themselves in differences in estimated flavonoid intake. These differences have ramifications for both clinical and population studies, in terms of ascertaining flavonoid exposure, and also for epidemiological association studies, where ranking of intake within a population is of importance. As such, when designing studies involving flavonoid intake assessment, it is important to carefully consider the different databases, and the methodologies underpinning them, in relation to the population and scientific question of interest.

ACKNOWLEDGEMENTS

The study was supported by research grants from Healthway Health Promotion Foundation of Western Australia, Sir Charles Gairdner Hospital Research Advisory Committee and by the project grants 254627, 303169 and 572604 from the National Health and Medical Research Council of Australia. The salary of KLI is supported by the National Health and Medical Research Council. None of these sources of support had any input into any aspect of the design and management of this study.

CONFLICT OF INTEREST STATEMENT

No authors declare a conflict of interest.

REFERENCES

1. Moss G, Smith P, Tavernier D. Glossary of class names of organic compounds and reactive intermediates based on structure. *International Union of Pure and Applied Chemistry* 1995;67(8):1307-75.

2. Aherne SA, O'Brien NM. Dietary flavonols: chemistry, food content, and metabolism. *Nutrition* 2002;18(1):75-81.

3. Cook N, Samman S. Flavonoids: chemistry, metabolism, cardioprotective effects, and dietary sources. *Journal of Nutritional Biochemistry* 1996;7(2):66-76.

4. Youdim KA, Martin A, Joseph JA. Incorporation of the elderberry anthocyanins by endothelial cells increases protection against oxidative stress. *Free Radical Biology and Medicine* 2000;29(1):51-60.

5. Hodgson J, Puddey I, Woodman R, Mulder T, Fuchs D, Scott K, et al. Effects of black tea on blood pressure: a randomized controlled trial. *Archives of Internal Medicine* 2012;17(2):186-88.

6. Bondonno CP, Yang X, Croft KD, Considine MJ, Ward NC, Rich L, et al. Flavonoid-rich apples and nitrate-rich spinach augment nitric oxide status and improve endothelial function in healthy men and women: a randomized controlled trial. *Free Radical Biology and Medicine* 2012;52(1):95-102.

7. Shrime MG, Bauer SR, McDonald AC, Chowdhury NH, Coltart CE, Ding EL. Flavonoid-rich cocoa consumption affects multiple cardiovascular risk factors in a meta-analysis of short-term studies. *The Journal of Nutrition* 2011;141(11):1982-88.

8. Hooper L, Kay C, Abdelhamid A, Kroon PA, Cohn JS, Rimm EB, et al. Effects of chocolate, cocoa, and flavan-3-ols on cardiovascular health: a systematic review and meta-analysis of randomized trials. *American Journal of Clinical Nutrition* 2012;95(3):740-51.

9. Shen Y, Ward NC, Hodgson JM, Puddey IB, Wang Y, Zhang D, et al. Dietary quercetin attenuates oxidant-induced endothelial dysfunction and atherosclerosis in apolipoprotein E knockout mice fed a high-fat diet: a critical role for heme oxygenase-1. *Free Radical Biology and Medicine* 2013;65:908-15.

10. Loke W, Proudfoot J, McKinley A, Hodgson J, Croft K. Pure dietary polyphenols attenuate atherosclerosis in apo E knockout mice via alleviating oxidative stress, inflammation and endothelial dysfunction. *Arteriosclerosis, Thrombos and Vascular Biology* 2010;30:749-57.

11. Shen Y, Croft KD, Hodgson JM, Kyle R, Lee I-LE, Wang Y, et al. Quercetin and its metabolites improve vessel function by inducing eNOS activity via phosphorylation of AMPK. *Biochemical Pharmacology* 2012;84(8):1036-44.

12. Huxley RR, Neil HA. The relation between dietary flavonol intake and coronary heart disease mortality: a meta-analysis of prospective cohort studies. *European Journal of Clinical Nutrition* 2003;57(8):904-8.

13. Hui C, Qi X, Qianyong Z, Xiaoli P, Jundong Z, Mantian M. Flavonoids, flavonoid subclasses and breast cancer risk: a meta-analysis of epidemiologic studies. *PLoS ONE* 2013;8(1):e54318.

14. Peters U, Poole C, Arab L. Does tea affect cardiovascular disease? a meta-analysis. *American Journal of Epidemiology* 2001;154(6):495-503.

15. Arab L, Liu W, Elashoff D. Green and black tea consumption and risk of stroke: a meta-analysis. *Stroke* 2009;40(5):1786-92.

16. Peterson JJ, Dwyer JT, Jacques PF, McCullough ML. Improving the estimation of flavonoid intake for study of health outcomes. *Nutrition reviews* 2015:nuv008.

17. US Department of Agriculture. USDA database for the flavonoid content of selected foods; release 2.1. Maryland, 2007.

18. Nutrient Data Laboratory, Agricultural Research Service, US Department of Agriculture. USDA database for proanthocyanidin content of selected foods. Beltsville, MD, 2004.

19. Mink PJ, Scrafford CG, Barraj LM, Harnack L, Hong C-P, Nettleton JA, et al. Flavonoid intake and cardiovascular disease mortality: a prospective study in postmenopausal women. *American Journal of Clinical Nutrition* 2007;85(3):895-909.

20. Chun OK, Chung SJ, Song WO. Estimated Dietary Flavonoid Intake and Major Food Sources of U.S. Adults. *The Journal of Nutrition* 2007;137(5):1244-52.

21. Neveu V, Perez-Jiménez J, Vos F, Crespy V, Du Chaffaut L, Mennen L, et al. Phenol-Explorer: an online comprehensive database on polyphenol contents in foods. *Database* 2010.

22. Ivey KL, Hodgson JM, Croft KD, Lewis JR, Prince RL. Flavonoid intake and all-cause mortality. *American Journal of Clinical Nutrition* 2015 101(5):1012-20.

23. Tresserra-Rimbau A, Rimm EB, Medina-Remón A, Martínez-González MA, De la Torre R, Corella D, et al. Inverse association between habitual polyphenol intake and incidence of cardiovascular events in the PREDIMED study. *Nutrition, Metabolism and Cardiovascular Diseases* 2014;24(6):639-47.

24. Jennings A, Welch AA, Fairweather-Tait SJ, Kay C, Minihane A-M, Chowienczyk P, et al. Higher anthocyanin intake is associated with lower arterial stiffness and central blood pressure in women. *The American Journal of Clinical Nutrition* 2012;96(4):781-88.

25. Service. USDoAAR. USDA database for the flavonoid content of selected foods, release 3.1, 2013.

26. Prince RL, Devine A, Dhaliwal SS, Dick IM. Effects of calcium supplementation on clinical fracture and bone structure: results of a 5-year, double-blind, placebo-controlled trial in elderly women. *Archives of Internal Medicine* 2006;166(8):869-75.

27. Hodge A, Patterson AJ, Brown WJ, Ireland P, Giles G. The Anti Cancer Council of Victoria FFQ: relative validity of nutrient intakes compared with weighed food records in young to middle-aged women in a study of iron supplementation. *Australia New Zealand Journal of Public Health* 2000;24(6):576-83.

28. Belobrajdic DP, Frystyk J, Jeyaratnaganthan N, Espelund U, Flyvbjerg A, Clifton PM, et al. Moderate energy restriction-induced weight loss affects circulating IGF levels independent of dietary composition. *European Journal of Endocrinology* 2010;162(6):1075-82.

29. Ivey KL, Lewis JR, Lim WH, Lim EM, Hodgson JM, Prince RL. Associations of proanthocyanidin intake with renal function and clinical outcomes in elderly women. *PLoS ONE* 2013;8(8):e71166.

30. Ireland P, Jolley D, Giles G, O'Dea K, Powles J, Ritishauser I, et al. Development of the Melbourne FFQ: a food frequency questionnaire for use in an Australian prospective study involving an ethnically diverse cohort. *Asia Pacific Journal of Clinical Nutrition* 1994;3:19-31.

31. Ivey KL, Lewis JR, Prince RL, Hodgson JM. Tea and non-tea flavonol intakes in relation to atherosclerotic vascular disease mortality in older women. *British Journal of Nutrition* 2013;110(9):1648-55.

32. Chun OK, Chung S-J, Claycombe KJ, Song WO. Serum C-Reactive Protein Concentrations Are Inversely Associated with Dietary Flavonoid Intake in U.S. Adults. *The Journal of Nutrition* 2008;138(4):753-60.

33. Walle T. Absorption and metabolism of flavonoids. *Free Radical Biology and Medicine* 2004;36(7):829-37.

34. Zamora-Ros R, Knaze V, Romieu I, Scalbert A, Slimani N, Clavel-Chapelon F, et al. Impact of thearubigins on the estimation of total dietary flavonoids in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *European Journal of Clinical Nutrition* 2013;67(7):779-82.

35. US Department of Agriculture. USDA database for the flavonoid content of selected foods release 1. Maryland, 2003.

36. Drossard C, Fröhling B, Dietrich H, Kersting M. Anthocyanin analysis in banana fruit—a mistake. *The American Journal of Clinical Nutrition* 2011;93(4):865-65.

37. Ambrosini GL, van Roosbroeck SAH, Mackerras D, Fritschi L, de Klerk NH, Musk AW. The reliability of ten-year dietary recall: implications for cancer research. *Journal of Nutrition* 2003;133(8):2663-68.

38. Åkerström A, Jaakola L, Bång U, Jäderlund A. Effects of latitude-related factors and geographical origin on anthocyanidin concentrations in fruits of Vaccinium myrtillus L.(bilberries). *Journal of Agricultural and Food Chemistry* 2010;58(22):11939-45.

39. McGhie TK, Hunt M, Barnett LE. Cultivar and growing region determine the antioxidant polyphenolic concentration and composition of apples grown in New Zealand. *Journal of Agricultural and Food Chemistry* 2005;53(8):3065-70.

40. Schindler M, Solar S, Sontag G. Phenolic compounds in tomatoes. Natural variations and effect of gamma-irradiation. *European Food Research and Technology* 2005;221(3-4):439-45.

Table 1: Structure and chemical name of flavonoid sub-classes included in this review.

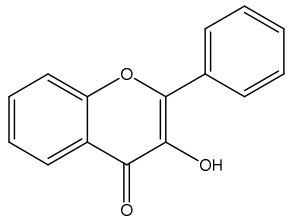
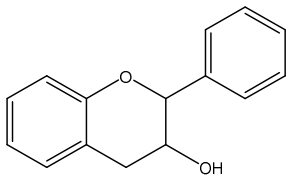
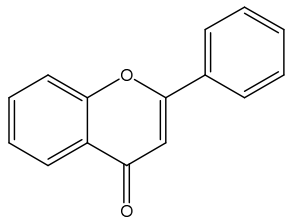
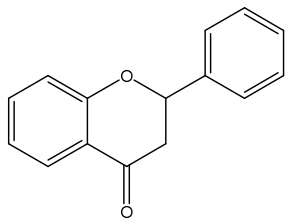
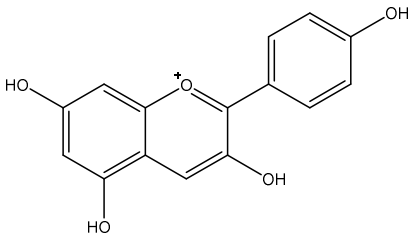
Sub-class	Chemical name	Characterising structure
Flavonol	3-hydroxy-2-phenylchromen-4-one	
Flavanol	3-hydroxy-2-phenylchroman	
Flavone	2-phenylchromen-4-one	
Flavanone	2-phenylchroman-4-one	
Anthocyanidin	4',3,5,7-hydroxy-2-phenylchromenylium	

Table 2: Daily consumption of total-flavonoid and flavonoid sub-class, as estimated with the United States Department of Agriculture and Phenol-Explorer databases

	Intake computed from USDA database (mg /d)	**Intake computed from PE database** (mg /d) [b]
Total-flavonoid [a]	834 ± 394	487 ± 243
Flavonoid sub-classes		
Flavonol [a]	30 ± 17	104 ± 61
Flavanol [a]	666 ± 345	327 ± 179
Flavone [a]	4 ± 3	13 ± 7
Flavanone [a]	40 ± 36	33 ± 31
Anthocyanidin [a]	88 ± 77	11 ± 11

Results are mean ± SD. [a] *Results are different by paired sample t-test (P<0.001).*

Table 3: Tertiles of Phenol-Explorer (PE) flavonoid intake expressed against United States Department of Agriculture (USDA) flavonoid intake tertiles

TOTAL-FLAVONOID [b]

	Total-flavonoid$_{PE}$ [c] Low	Moderate	High	
Total-flavonoid$_{USDA}$ [d]				
Low intake	315	38	1	**Classified identically:** 911 (86%)
Moderate intake	36	281	38	**Classified differently:** 152 (14%)
High intake	3	36	315	**Correlation coefficient:** 0.884 [a]

FLAVONOL [b]

	Flavonol$_{PE}$ [e] Low	Moderate	High	
Flavonol$_{USDA}$ [f]				
Low intake	289	65	0	**Classified identically:** 789 (74%)
Moderate intake	49	226	80	**Classified differently:** 274 (26%)
High intake	16	64	274	**Correlation coefficient:** 0.773 [a]

FLAVANOL [b]

	Flavanol$_{PE}$ [g] Low	Moderate	High	
Flavanol$_{USDA}$ [h]				
Low intake	303	49	2	**Classified identically:** 867 (82%)
Moderate intake	45	261	49	**Classified differently:** 196 (18%)
High intake	6	45	303	**Correlation coefficient:** 0.845 [a]

FLAVONE [b]

	Flavone$_{PE}$ [i] Low	Moderate	High	
Flavone$_{USDA}$ [j]				
Low intake	184	111	59	**Classified identically:** 483 (45%)
Moderate intake	91	134	130	**Classified differently:** 580 (55%)
High intake	79	110	165	**Correlation coefficient:** 0.298 [a]

FLAVANONE [b]

	Flavanone$_{PE}$ [k] Low	Moderate	High	
Flavanone$_{USDA}$ [l]				
Low intake	347	7	0	**Classified identically:** 1033 (97%)
Moderate intake	7	340	8	**Classified differently:** 30 (3%)
High intake	0	8	346	**Correlation coefficient:** 0.979 [a]

ANTHOCYANIDIN [b]

	Anthocyanidin$_{PE}$ [m] Low	Moderate	High	
Anthocyanidin$_{USDA}$ [n]				
Low intake	259	78	17	**Classified identically:** 729 (69%)
Moderate intake	82	203	70	**Classified differently:** 334 (31%)
High intake	13	74	267	**Correlation coefficient:** 0.701 [a]

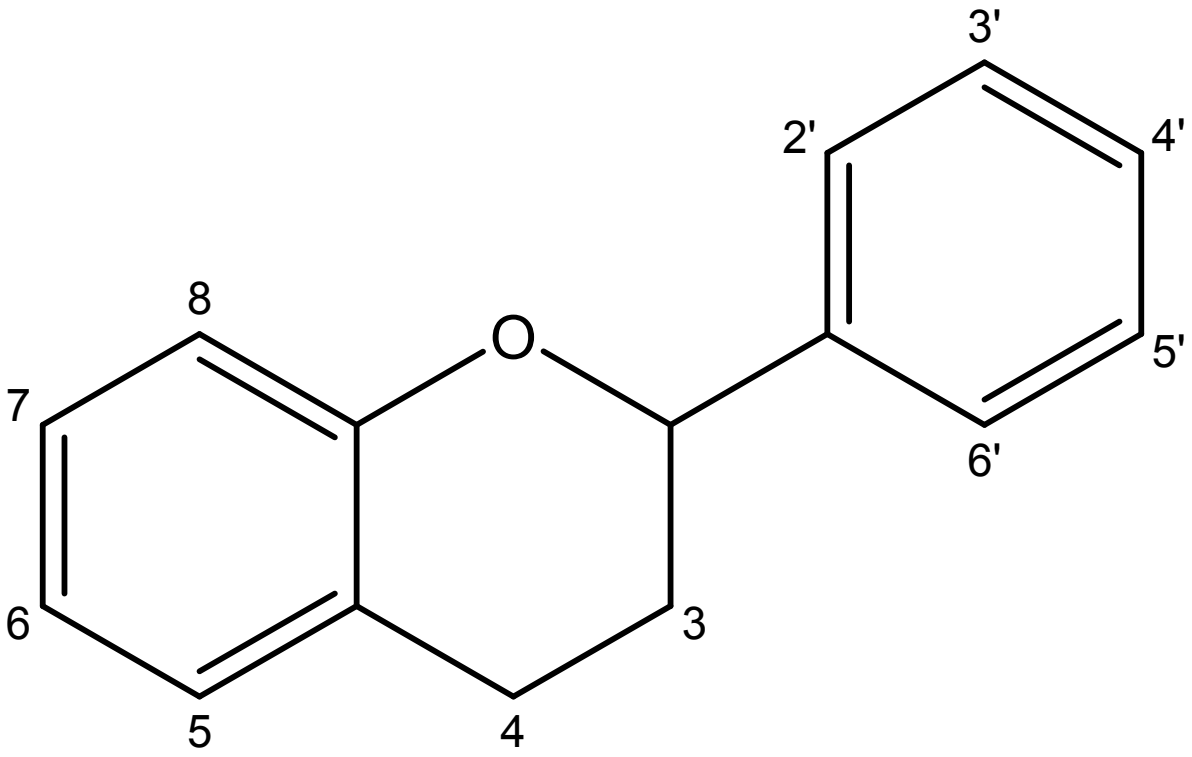
Results are n, n (%), or Spearman rank correlation coefficient, where appropriate.

[a] *Results are significantly correlated (P<0.001).*

[b] *Results are significantly different by Pearson Chi-Square (P<0.001).*

[c] *$Flavonoid_{PE}$:*	*Low (<371mg/d);*	*Moderate (371-<573 mg/d);*	*High (≥573mg/d).*
[d] *$Flavonoid_{USDA}$:*	*Low (<646mg/d);*	*Moderate (646-<976mg/d);*	*High (≥976mg/d).*
[e] *$Flavonol_{PE}$:*	*Low (<74mg/d);*	*Moderate (74-<130mg/d);*	*High (≥130mg/d).*
[f] *$Flavonol_{USDA}$:*	*Low (<27mg/d);*	*Moderate (27-<40mg/d);*	*High (≥40mg/d).*
[g] *$Flavanol_{PE}$:*	*Low (<241mg/d);*	*Moderate (241-<370mg/d);*	*High (≥370 mg/d).*
[h] *$Flavanol_{USDA}$:*	*Low (<502mg/d);*	*Moderate (502-<791mg/d);*	*High (≥791mg/d).*
[i] *$Flavone_{PE}$:*	*Low (<10mg/d);*	*Moderate (10-<15mg/d);*	*High (≥15 mg/d).*
[j] *$Flavone_{USDA}$:*	*Low (<3mg/d);*	*Moderate (3-<5mg/d);*	*High (≥5mg/d).*
[k] *$Flavanone_{PE}$:*	*Low (<14mg/d);*	*Moderate (14-<40mg/d);*	*High (≥40 mg/d).*
[l] *$Flavanone_{USDA}$:*	*Low (<18mg/d);*	*Moderate (18-<48mg/d);*	*High (≥48mg/d).*
[m] *$Anthocyanin_{PE}$:*	*Low (<4mg/d);*	*Moderate (4-<11mg/d);*	*High (≥11 mg/d).*
[n] *$Anthocyanin_{USDA}$:*	*Low (<40mg/d);*	*Moderate (40-<98mg/d);*	*High (≥98mg/d).*

Figure 1: chemical structure of the flavan (2-phenylchroman) nucleus of flavonoid molecules



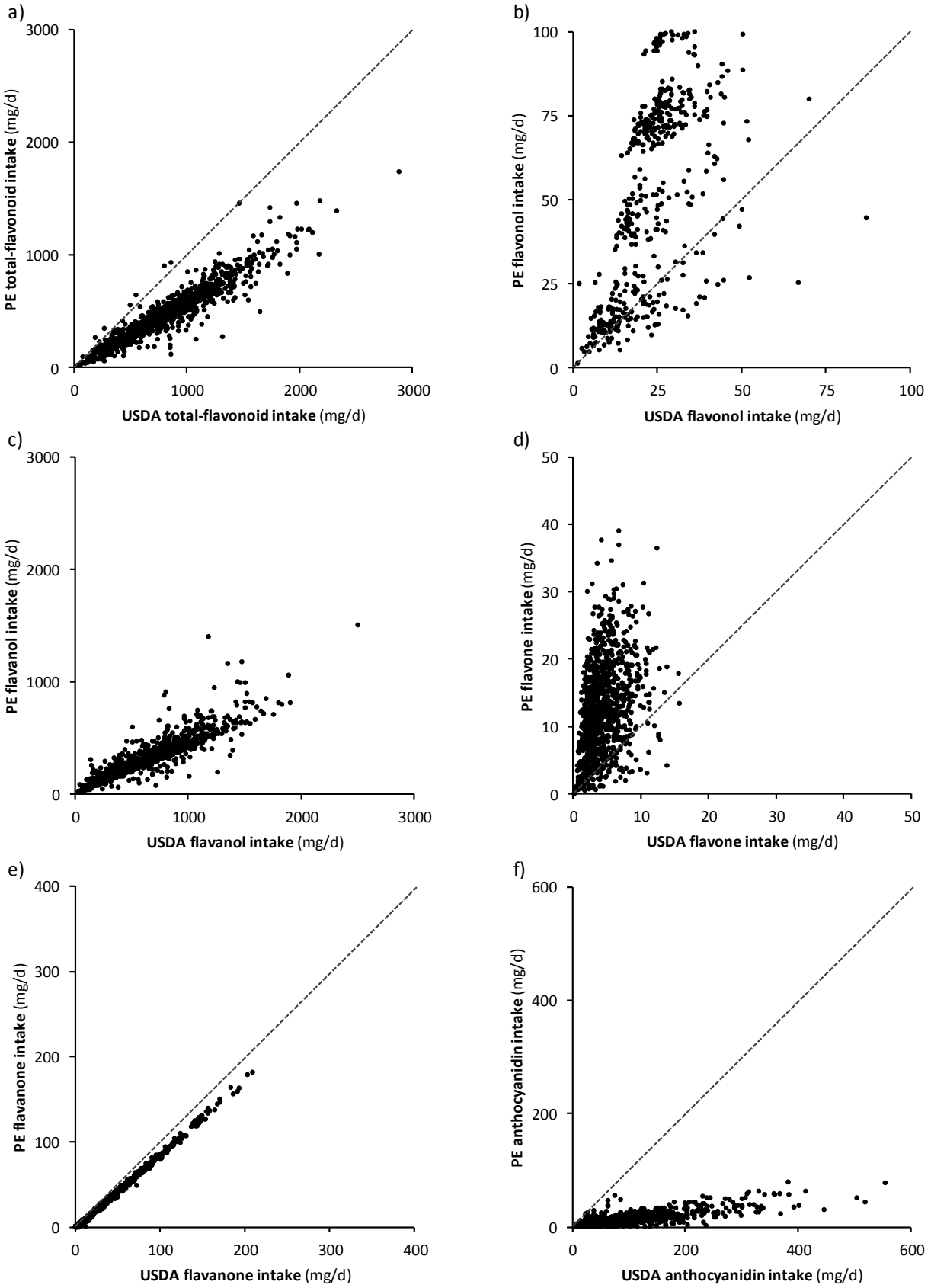


Figure 2: Level of agreement between United States Department of Agriculture (USDA) and Phenol-Explorer (PE) total-flavonoid and flavonoid sub-class intake estimates.

n = 1,063.----- : line of equality.

a) Total-flavonoid intake estimates

Pearson correlation coefficient= 0.943, P<0.001.

Unstandardized B= 0.581 ± 0.006, P<0.001.

b) Flavonol intake estimates

Pearson correlation coefficient= 0.830, P<0.001.

Unstandardized B= 3.076 ± 0.063, P<0.001.

c) Flavanol intake estimates

Pearson correlation coefficient= 0.892, P<0.001.

Unstandardized B= 0.461 ± 0.007, P < 0.001.

d) Flavone intake estimates

Pearson correlation coefficient= 0.340, P<0.001.

Unstandardized B= 0.890 ± 0.075, P<0.001.

e) Flavanone intake estimates

Pearson correlation coefficient= 0.999, P<0.001.

Unstandardized B= 0.857 ± 0.001, P<0.001.

f) Anthocyanidin intake estimates

Pearson correlation coefficient= 0.817, P<0.001.

Unstandardized B= 0.120±0.003, P<0.001.

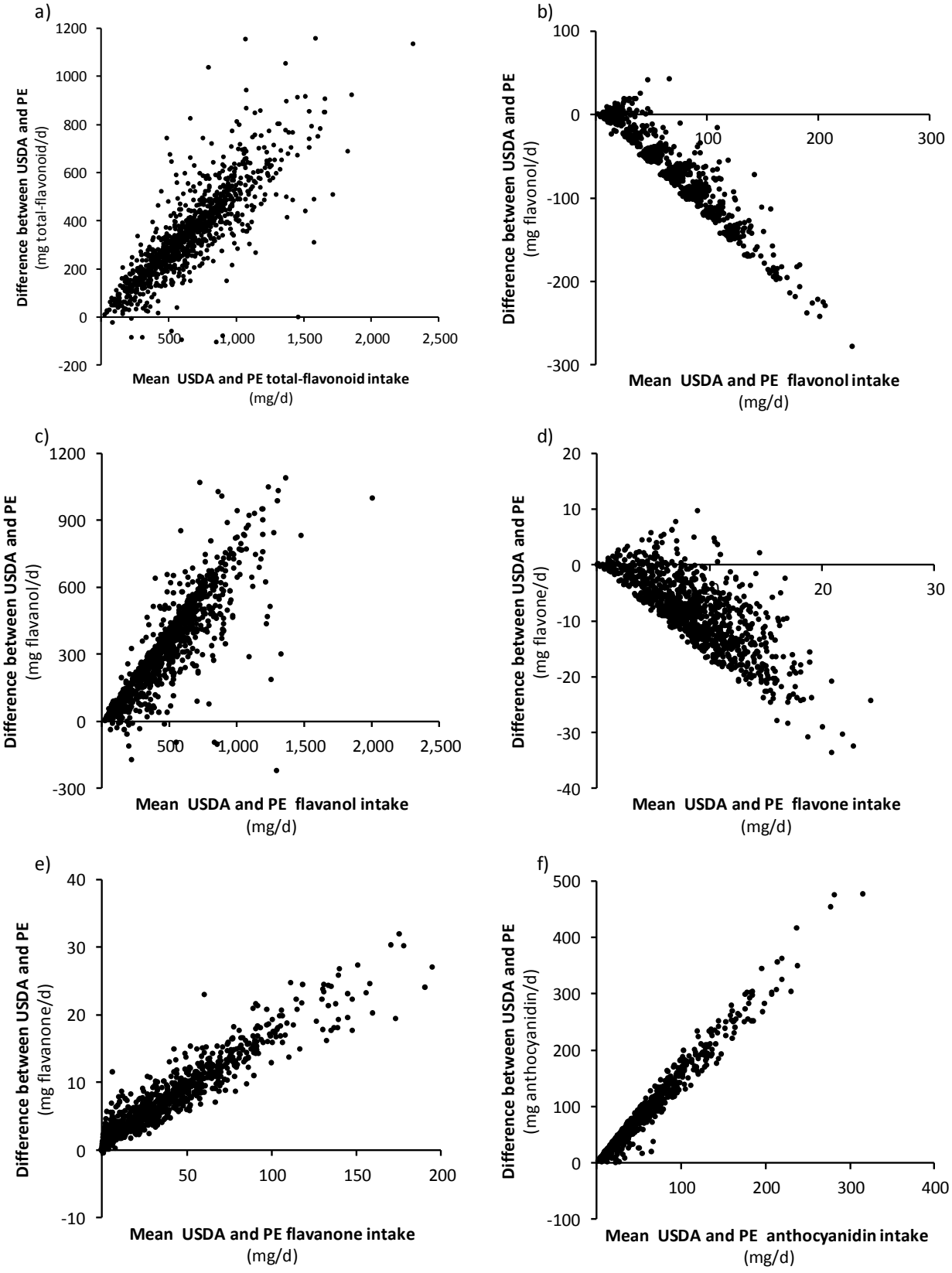


Figure 3: Relationship between level of disagreement and mean estimated total-flavonoid and flavonoid sub-class intake

n = 1,063.

a) Bland and Altman plot incorporating total-flavonoid intake estimates

b) Bland and Altman plot incorporating Flavonol intake estimates

c) Bland and Altman plot incorporating Flavanol intake estimates

d) Bland and Altman plot incorporating Flavone intake estimates

e) Bland and Altman plot incorporating Flavanone intake estimates

f) Bland and Altman plot incorporating Anthocyanidin intake estimates