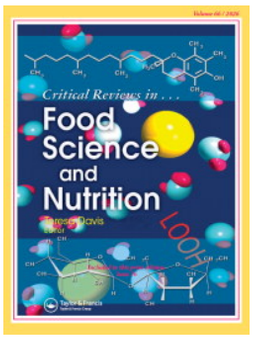

Critical Reviews in Food Science and Nutrition

ISSN: 1040-8398 (Print) 1549-7852 (Online) Journal homepage: www.tandfonline.com/journals/bfsn20

A scoring system for validity assessment of biomarkers of food intake

Christoph Hassenberg, Carina I. Mack, Claudia Favari, Sylvie Baier, Ramon Estruch, Gabi Kastenmüller, Rosa Lamuela-Raventós, Rikard Landberg, Pedro Mena, Costanza Michelini, Stefania Noerman, Ana Rodriguez-Mateos, Mia Stråvik, Francisco A. Tomás-Barberán, Lara Vehovec, David S. Wishart, Claudine Manach & Sabine E. Kulling

To cite this article: Christoph Hassenberg, Carina I. Mack, Claudia Favari, Sylvie Baier, Ramon Estruch, Gabi Kastenmüller, Rosa Lamuela-Raventós, Rikard Landberg, Pedro Mena, Costanza Michelini, Stefania Noerman, Ana Rodriguez-Mateos, Mia Stråvik, Francisco A. Tomás-Barberán, Lara Vehovec, David S. Wishart, Claudine Manach & Sabine E. Kulling (29 Apr 2026): A scoring system for validity assessment of biomarkers of food intake, Critical Reviews in Food Science and Nutrition, DOI: 10.1080/10408398.2026.2633561

To link to this article: https://doi.org/10.1080/10408398.2026.2633561

CRITICAL REVIEWS IN FOOD SCIENCE AND NUTRITION
https://doi.org/10.1080/10408398.2026.2633561

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS Check for updates

A scoring system for validity assessment of biomarkers of food intake

Christoph Hassenberg[a*], Carina I. Mack[a*], Claudia Favari[b], Sylvie Baier[c], Ramon Estruch[d,e], Gabi Kastenmüller[c], Rosa Lamuela-Raventós[e,f], Rikard Landberg[g], Pedro Mena[b], Costanza Michelini[b], Stefania Noerman[g#], Ana Rodriguez-Mateos[h], Mia Stråvik[g], Francisco A. Tomás-Barberán[i], Lara Vehovec[c], David S. Wishart[j], Claudine Manach[k**] and Sabine E. Kulling[a**]

[a]Department of Safety and Quality of Fruit and Vegetables, Max Rubner-Institut, Karlsruhe, Germany; [b]Human Nutrition Unit, Department of Food and Drugs, University of Parma, Parma, Italy; [c]Institute of Computational Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; [d]Hospital Clínic, University of Barcelona, Cardiovascular Risk, Nutrition and Aging Research Group, Barcelona, Spain; [e]Spanish Biomedical Research Centre in Physiopathology of Obesity and Nutrition (CiberOBN), Madrid, Spain; [f]Department of Nutrition, Food Science and Gastronomy, University of Barcelona, Barcelona, Spain; [g]Division of Food and Nutrition Science, Department of Life Sciences, Chalmers University of Technology, Gothenburg, Sweden; [h]Department of Nutritional Sciences, School of Life Course and Population Sciences, Faculty of Life Sciences and Medicine, King's College London, London, UK; [i]Department of Food Science, CEBAS-CSIC, Spanish Research Council, Murcia, Spain; [j]Departments of Computing Science and Biological Sciences, University of Alberta, Edmonton, Canada; [k]Université Clermont Auvergne, INRAE, Human Nutrition Unit, Clermont-Ferrand, France

ABSTRACT

Recently, numerous novel putative biomarkers of food intake (BFIs) have been discovered to complement self-reported dietary assessment. However, few BFIs have been properly validated for their intended applications due to lack of comprehensive validation frameworks, need for complementary expertise and scattered data in the literature. Although a few biomarker validation frameworks have been proposed, they still require improvement to fully encompass relevant criteria and provide precise guidance for harmonized evaluation of BFI validity. The FoodPhyt consortium developed a new scoring system based on the FoodBAll validation scheme, to standardize scoring and allow comparison of BFIs with varying validity. The following criteria were included: level of identification and plausibility (pass-or-fail criteria), specificity, variability in plants/foods and biological samples, dose-response, robustness, and analytical characterization. The score is bipartite and includes availability of supporting literature (data availability), and BFI quality (performance). Data availability helps assessing confidence in the assigned score, while highlighting knowledge gaps. Detailed guidelines and examples are provided to facilitate BFI scoring and reproducible application. An online repository, BFI-Hub (https://biomarker.plantintake.eu/app/), was developed to share validity scores of BFIs and detailed validation data. This database represents a key step to make BFI information universally accessible and facilitate application of BFIs in practice.

KEYWORDS

Biomarker of food intake; biomarker validation; dietary assessment; online repository; practical guidance; scoring system

General introduction

Dietary assessment traditionally relies on self-reporting methods such as food frequency questionnaires (FFQ), dietary recalls, and food diaries. These approaches have well-known limitations due to their subjective nature and incompleteness of associated food composition tables (Brennan 2017; Garcia-Aloy et al. 2017). Consequently, the associations between diet/dietary components and health outcomes in epidemiological studies may be prone to random and systematic errors, which may result in attenuated or inflated associations (Subar et al. 2015; Brennan 2017; Ottaviani et al. 2023).

Biomarkers of food intake (BFIs), which serve as objective measures of dietary intake, offer a promising approach to improve the precision of dietary assessment, when used as a complement to traditional methods (Jenab et al. 2009; Brennan 2017; Garcia-Aloy et al. 2017). BFIs can also be used to assess compliance in intervention studies. In recent years, significant progress has been made in discovering several hundred new BFI candidates using metabolomics (Rothwell et al. 2014; Cabezas et al. 2016; Zhu et al. 2016; Saenger et al. 2017; Vazquez-Manjarrez et al. 2019; Beckmann et al. 2020; Garcia-Aloy et al. 2020; McNamara et al. 2020; Tomás-Navarro et al. 2021; Landberg et al. 2023; Cuparencu et al. 2024; Hassenberg et al. 2026). However, robust validation and translation into practice has been achieved for only few of them, with the majority of BFI candidates only partly validated at best (Dragsted et al. 2018; D'Angelo et al. 2019; Unión-Caballero et al. 2024). Moreover, BFI validation data are dispersed throughout the literature, making it challenging for any researchers, both to select appropriate BFIs, and judge their validity and applicability.

CONTACT Carina I. Mack carina.mack@mri.bund.de
Supplemental data for this article can be accessed online at https://doi.org/10.1080/10408398.2026.2633561.
*These authors are co-first authors to this work.
**These authors are co-senior authors to this work.
#Clinical Nutrition Research Centre (CNRC), Singapore Institute of Food and Biotechnology Innovation (SIFBI), Agency for Science, Technology and Research (A*STAR), 14 Medical Drive, MD6, #07-02, Singapore 117599, Singapore.

Several review papers on BFIs for specific foods or food groups including their validation data have been published (Michielsen et al. 2018; Munger et al. 2018; Praticò et al. 2018; Rothwell et al. 2018; Sri Harsha et al. 2018; Ulaszewska et al. 2018; Cuparencu et al. 2019; Garcia-Aloy et al. 2019; Landberg et al. 2019; Vazquez-Fresno et al. 2019; Xi and Dragsted 2019; Zhou et al. 2019; Brouwer-Brolsma et al. 2020; Clarke et al. 2020; Ulaszewska et al. 2020; Vazquez-Manjarrez et al. 2020; Li et al. 2021; de la O et al. 2025). Additionally, databases, such as ExposomeExplorer (http://exposome-explorer.iarc.fr/) and MarkerDB (https://markerdb.ca/), provide useful information on various biomarkers (e.g., types of biomarkers, structures, concentrations), but have only a partial coverage of BFIs and do not inform on their validation status. Importantly, the Joint Programming Initiative "A healthy diet for a healthy life" (JPI HDHL) FoodBAll (*Food Biomarkers Alliance*) project developed a consensus-based qualitative procedure for assessing the current validation status of BFIs using an unbiased and systematic approach (Dragsted et al. 2018). The FoodBAll scheme included eight validation criteria: plausibility, dose-response, time-response, robustness, reliability, stability, analytical performance, and reproducibility, and marked an important step toward standardizing BFI validation (Dragsted et al. 2018). Nonetheless, the FoodBAll validation scheme was primarily intended to help identify knowledge gaps for full BFI validation rather than to serve as a scoring system. As such, the scheme offers only "yes/no/unknown" options if a BFI matches all requirements or not, with no possibility for gradual rating. Furthermore, the evaluation of the FoodBAll validation criteria can be subjective, as there are no defined cutoffs or boundaries for meeting certain criteria.

Rafiq et al. (2021) proposed a simple scoring system based on the number of studies reporting a BFI, with 2 points assigned for each interventional study and an additional point for each observational study. However, a BFI can be repeatedly associated with the intake of a food in several studies without being of interest due to a lack of specificity, as exemplified by hippuric acid (Rafiq et al. 2021). More recently, Cuparencu et al. (2024) proposed a four-level utility scale to rank BFIs according to the information available for the FoodBAll validation criteria; plausibility, robustness and reliability. This new scale offers the advantage of providing a straightforward selection of the most validated BFIs (utility level 1), while distinguishing between BFIs that are plausible and robust, but still require validation for reliability (utility level 2), and those BFIs that appear plausible but have only been reported in one type of study (intervention or observational) and thus not yet confirmed to be robust or reliable (utility level 3). While useful to classify BFIs by their overall quality and utility, this scheme does not provide detailed information on how BFIs within the same utility level perform with respect to individual validation criteria.

Recognizing the need for a universal reference system to rate the validity of unambiguous BFIs while ensuring traceability, we developed the FoodPhyt scoring system based on previous validation frameworks. The scoring system was initially developed using BFIs for plant-based foods, but it should also be applicable to animal-derived foods. The online database BFI-Hub (https://biomarker.plantintake.eu/app/) was developed to share the scoring and validation data of BFIs enabling transparent traceability. BFI-Hub facilitates an informed choice of the best suitable BFIs for a given application in human studies based on a BFIs quality (performance; ability to fulfill the validation criteria) and data availability. The manuscript describes the development of the scoring system and details how the validation criteria were considered to assess BFIs validity (see respective section for each validation criterion).

Methodology to develop the scoring system

All procedural steps are depicted in the accompanying flowchart (Figure 1). The scoring system was developed by the FoodPhyt consortium through consensus, building on the eight FoodBAll validation criteria (Dragsted et al. 2018). We extracted literature data related to the FoodBall criteria for >500 BFIs reflecting the 30 most-consumed plant foods in Europe (Supplemental Material I) and assessed which criteria could be applied in practice. The "reliability" criterion proposed by Dragsted et al. (2018) was particularly difficult to apply since no gold standard dietary assessment instrument or BFI yet exists without any associated bias. The comprehensive analysis of the collected data revealed gaps in the current BFI validation assessment that led to the addition of two new criteria: level of identification and variability (related to humans and food sources). A second literature search was then conducted accordingly. Marked variability in supporting data across BFIs prompted the inclusion of data availability as one part of the global score to capture strength of available evidence. Finally, BFI validity was considered as the combination of BFI quality and data availability.

A set of 27 BFI-food pairs spanning the full spectrum (poor to excellent, single foods *vs.* food groups, high *vs.* low data availability) was used to test the existing and new validation criteria. Numerical thresholds and weightings were iteratively refined until the resulting score aligned with the expert consensus reached by specialists in dietary biomarkers, nutrition, food chemistry, metabolism and metabolomics from the FoodPhyt consortium. Details on optimization of weighting and numeric thresholds can be found in Supplemental Material I. In brief, thresholds for each validation criterion were chosen by examining the distribution of the underlying data: (i) any negative values arising from penalization indicating poorest quality (IV), with the remaining numeric range being divided into equal intervals, (ii) visual inspection of the empirical score distribution revealing natural gaps/minima, and (iii) those gaps best separate high- from low-quality BFIs and fit with the expert consensus opinion collected within the FoodPhyt consortium. Individual criterion sub-scores were then combined using consensus-derived weights and transformed into categorical levels of data availability and overall BFI quality. A one-factor-at-a-time sensitivity analysis was performed to test how variations in weights and thresholds affect the rankings (Supplemental Figures S1–S4). The robustness of the applied

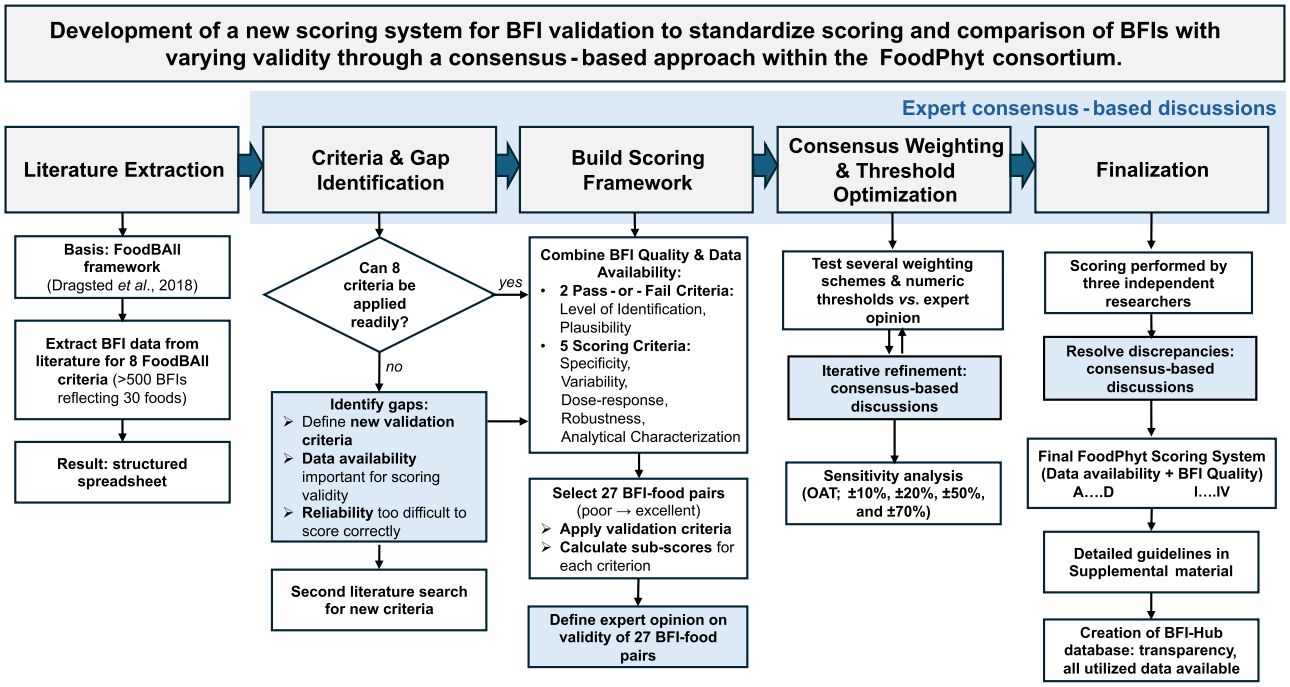

Figure 1. Schematic workflow for the consensus-based development of a BFI validation scoring system within the FoodPhyt consortium. The diagram follows five sequential modules. First, literature extraction gathers validation data for >500 BFIs using the original FoodBAll criteria. Second, criteria & gap identification evaluates the applicability of those criteria; a diamond-shaped decision symbol marks the central question posed during the literature review. Third, a provisional scoring framework is built by integrating the original and newly identified criteria, incorporating data-availability, and selecting 27 BFI-food pairs for application and testing. Fourth, expert consensus weighting and threshold optimization (including sensitivity analyses) refines numeric thresholds and relative weights and aligns calculated scores with expert opinion. Finally, the system is finalized through independent scoring by three researchers and resolution of any resulting discrepancies. Steps that required expert-consensus discussions are highlighted in light-blue boxes, and arrows indicate the direction of information flow throughout the process. OAT: one-factor-at-a-time.

calculation method was confirmed, as moderate changes in weighting and thresholds did not affect the scoring results.

Finally, to ensure consistency and repeatability of the scoring system, three independent researchers applied the scoring guidelines to the selected BFI-food pairs in an iterative process, compared results among themselves and with the expert consensus opinion. Discrepancies were discussed and used to refine the scoring guidelines, calculations, and weightings.

Validation criteria and the scoring system

A BFI should be an unambiguously identified molecule present in biological samples after intake of a certain food and the validity of a BFI should always be evaluated for a clearly defined food or food group.

The FoodPhyt score is bipartite and consists of two parts that together provide information on the validity of the BFI:

- Data Availability: This considers whether the underlying data is adequate to assess the BFI quality, rated in four levels from A (highest availability) to D (lowest availability).
- BFI Quality: This describes the biomarker's ability to fulfill the validation criteria, and therefore, accurately reflect the consumption of a specific food or food group, rated in four levels from I (highest quality) to IV (lowest quality).

The overall scheme of the scoring system is described in Figure 2. The criteria included in the scoring system to judge the overall biomarker validity were: level of identification, plausibility, specificity, variability, dose-response, robustness, and analytical characterization. The level of identification (molecule identity) and plausibility of a BFI were implemented as pass-or-fail criteria. The other criteria are part of a flow chart used to calculate sub-scores for each criterion. Details on the criteria and their sub-score calculations are provided in the following sections, along with examples. An illustrated step-by-step example (phloretin 2′-O-glucuronide) is provided in the Supplemental Material I, including Supplemental Table S1 and Figures S5–S10. For the final score calculation, the relative points for each sub-score are weighted, summed and translated to the four levels separately for BFI quality and data availability (Figure 2). The five sub-scores for specificity, variability, dose-response, robustness and analytical characterization are weighted according to their importance: 3, 3, 3, 3, and 1, respectively. Analytical characterization received a lower weighting since it can be improved through further method development or reference standard synthesis. The results obtained for a selection of 27 BFI-food pairs are detailed in Table 1 to illustrate the evaluation process for each validation criterion. The underlying validation data for the examples are listed together with the corresponding references in the Supplemental Material II Tables S2–S4 and the scoring in Supplemental Table S5.

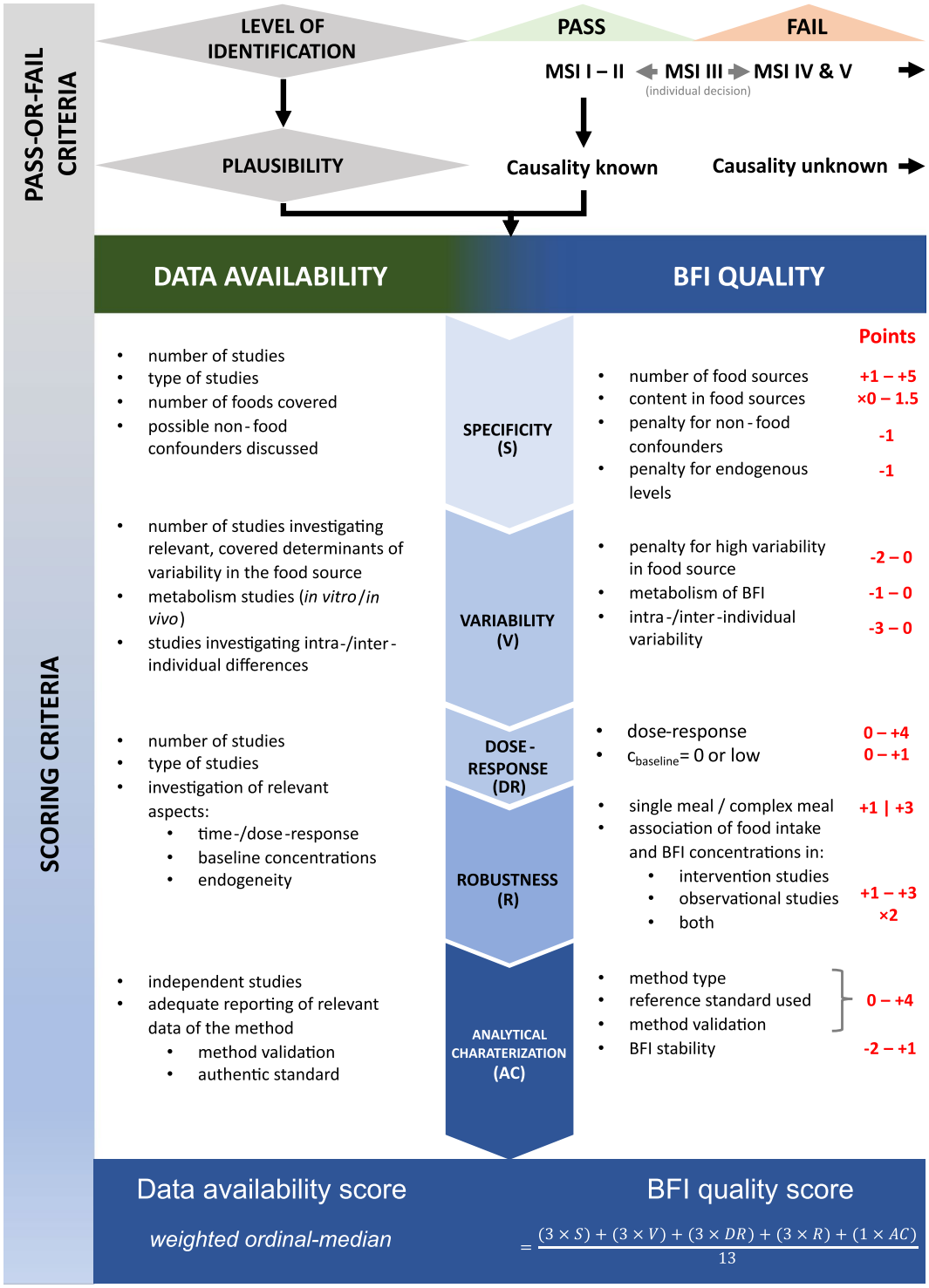

Figure 2. Overall scheme of the scoring system showing the pass-or-fail criteria (level of identification, plausibility) and the scoring criteria (specificity, variability, dose-response, robustness and analytical characterization) that are individually evaluated to score a BFI. The pass-or-fail criteria are the minimum requirements to further score BFIs. The identification confidence level – according to the MSI guidelines (2014) (Schymanski et al 2014) – has to be at least Level 2 and the origin of the BFI must be known. When the pass-or-fail criteria are fulfilled the scoring of the BFI is performed in the order of the five validation criteria shown in the middle. On the left side the corresponding aspects for evaluating data availability are presented. On the right side the aspects and possible points for the corresponding validation criterion are briefly summarized. Once the individual sub-scores for the scoring criteria (BFI quality and data availability) have been determined, the overall score for the BFI is calculated.

For the scoring, all relevant literature must be comprehensively collected and different types of publications might be needed. While primary research articles are preferred for evaluating the quality of a BFI, high-quality review articles may serve as suitable sources of information for well-established aspects of a biomarker or to provide an overview, e.g., the presence of the BFI in other foods.

Level of identification criterion

The level of identification of the molecule is a new criterion, and was added to avoid the propagation of falsely or insufficiently identified putative BFIs. The level of identification of a BFI is defined by the unambiguous description of its chemical structure. The International Union of

Pure and Applied Chemistry (IUPAC) nomenclature or the machine-readable International Chemical Identifier (InChI) or InChIKey are recommended (Heller et al. 2013). Common names and synonyms sometimes lack precision, which in the worst case can lead to misinterpretation. Furthermore, in any metabolomic study, the correct identification of detected features remains challenging, depending on classification by authors and presence in databases or availability of analytical standards. In 2007, the Chemical Analysis Working Group (CAWG) of the Metabolomics Standards Initiative (MSI) proposed a four-level framework to report metabolite identification from confirmed identifications (level 1) to unknown features (level 4) (Sumner et al. 2007). This system was widely accepted in the community (Creek et al. 2014) and refined as a five-level system in 2014 (Schymanski et al. 2014). The level of identification based on this consensus system was included as the first pass-or-fail criterion in the BFI scoring system (see Figure 2).

To pass the criterion, a BFI must be identified at MSI Level 1 or 2 (2007 guidelines) or MSI Level 1–3 (2014 guidelines) in at least one publication. When the level of identification criterion is considered as failed, the feature should not be scored for validity. Nevertheless, the analytical information available in the original publication remains important, as it allows other laboratories to recognize the feature and potentially complete its identification, especially if it repeatedly appears in association with the consumption of specific foods across studies. Once unambiguously identified, the putative BFI can be scored.

The following examples show BFIs that passed or failed the level of identification criterion:

- Phloretin 2'-*O*-glucuronide, a urinary BFI candidate for apple intake, achieved MSI Level 1 through comparison with an in-house synthesized reference standard in two studies (Marks et al. 2009; Trost et al. 2018).
- Ergothioneine, an amino acid derivative and putative BFI for mushroom intake, was identified by comparison of its MS^1 and MS^2 spectra with entries in a public database. Since no reference standard was used (Pallister et al. 2016; Wang et al. 2018), the identification corresponds to MSI Level 2 (according 2014 guidelines) and therefore, the marker passes the level of identification criterion.
- "4-guanidinobutanoic acid with an isoprene modification" was found associated with spinach intake in an intervention study. The identification was inferred from precursor mass and the interpretation of MS^2 spectra. The fragmentation pattern provided some supporting evidence for the presence of the 4-guanidinobutanoic acid moiety, while the presence of the isoprene unit remains speculative due to the lack of reference spectra in databases (Lynn et al. 2019). According to the MSI 2014 guidelines, this corresponds to MSI Level 4, and since the identification is not unambiguous, the criterion is failed.

Plausibility criterion

The plausibility of a BFI, also considered in the FoodBAll scheme, represents the causal link between an ingested food or food group, that is clearly defined, and the BFI presence in biological samples. It is a crucial validation criterion and therefore implemented as a pass-or-fail criterion in the FoodPhyt scoring system. A simple correlation in an observational study between a metabolite and food intake data is not sufficient. To grant plausibility of a BFI, it is essential to understand how a BFI arises in biological samples and originates from food consumption, either as a native compound, a product generated during food processing, or a specific metabolite produced by the body. For example, ergothioneine is synthesized in mushrooms (Genghof 1970), which explains its association with mushroom intake in observational studies (Pallister et al. 2016; Wang et al. 2018; 2020).

Plausibility is typically shown by detecting the BFI after consumption of a particular food in controlled intervention studies and assessing its origin. If the BFI is a metabolite derived from a precursor compound in the food, the metabolic reactions involved, whether mediated by host enzymes (phase I or II), gut microbiota, or both, should be documented. The complete set of metabolites formed from the precursor may also be of interest, to reveal alternative BFI candidates, in case analytical challenges arise with the initially selected BFI.

To perform the scoring for the plausibility criterion, the systematic literature search (search strategy see Supplemental Material I) should: (i) use scientific and common names of the food, and (ii) include common synonyms of BFIs in the search. Peer-reviewed original articles from at least two independent research groups or review articles are recommended to confirm the occurrence of a BFI or its precursor in a given food, while databases should only be used as a starting point since they may contain errors. Various types of publication are useful to demonstrate plausibility: (i) analysis of the BFI or its precursor(s) in the target food, (ii) description of plant-specific biosynthetic pathways for the BFI, (iii) the chemical formation route of a BFI during food processing, and/or (iv) the biotransformation pathway of a precursor in humans (i.e., phase I, phase II, intestinal microbiota metabolism, or non-enzymatic reaction). Human intervention studies are considered in priority, but *in vitro* models specifically aiming to investigate human metabolic reactions may also be used to document the metabolism. Of note, the collected information is relevant for both BFI plausibility and variability (see variability criterion). While plausibility addresses the mechanistic link between food and BFI, variability evaluates how consistently the BFI reflects intake across individuals.

The following examples illustrate some key aspects necessary for assessing plausibility, as well as potential pitfalls:

- Phloretin 2'-*O*-glucuronide, a BFI for apple intake, originates from human/host co-metabolism of phloretin glucosides that are precursors present in apples, which are cleaved by the gut microbiota into phloretin and subsequently metabolized by phase II enzymes into various glucuronide, sulfate, and sulfo-glucuronide

Table 1. Results from the evaluation of the validation data (provided in Supplemental Tables 2–4) for the selected BFIs (Part I including the criteria: level of identification and plausibility).

			Identification & Plausibility			
BFI name	Associated food	MSI Level (2014)	Mediator of BFI metabolism	BFI was reported in biological samples, and the BFIs origin is known	Is the metabolism of the BFI known?	Fail-or-pass
11β,13-Dihydrolactucin	Lettuce	1	UC	yes	yes	pass
Arbutin	Pear	3	UC	yes	yes	pass
N-Caproylhistamine	Tomato	1	UC	yes	partly	pass
Salsolinol sulfate	Banana	1	HM	yes	yes	pass
Phloretin*	Apple**	1	HM	yes	yes	pass
Phloretin 2-*O*-glucuronide	Apple**	1	HM	yes	yes	pass
Ergothioneine	Mushrooms	2	UC	yes	partly	pass
Ergothioneine	Champignon	2	UC	yes	partly	pass
Sulforaphane	Brassicaceae	1	HMGM	yes	yes	pass
Sulforaphane-*N*-acetylcysteine	Brassicaceae	1	HM	yes	yes	pass
Sulforaphane-*N*-acetylcysteine	Broccoli	1	HM	yes	yes	pass
3,5-Dihydroxybenzoic acid	Whole Grain	1	HM	yes	yes	pass
3,5-Dihydroxybenzoic acid	Rye	1	HM	yes	yes	pass
3,5-Dihydroxybenzoic acid	Wheat	1	HM	yes	yes	pass
Atractyligenin glucuronide	Coffee	1	HM	yes	yes	pass
Cyclo(L-leu-L-pro)	Coffee	1	UC	yes	no	pass
Caffeine	Coffee	1	UC	yes	yes	pass
1-Methylxanthine	Coffee	1	HM	yes	yes	pass
Trigonelline	Coffee	1	UC	yes	yes	pass
Trigonelline	Green Beans	1	UC	yes	yes	pass
Trigonelline	Peas	1	UC	yes	yes	pass
Hesperetin	Orange	1	GM	yes	yes	pass
Hesperetin 3-*O*-glucuronide	Orange	1	HMGM	yes	yes	pass
Hesperetin 7-*O*-glucuronide	Orange	1	HMGM	yes	yes	pass
Proline betaine	Orange	1	UC	yes	partly	pass
Proline betaine	Citrus fruits	1	UC	yes	partly	pass
Valencic acid glucuronide	Celery	3	HM	no	no	**fail**

		Specificity							
BFI name	Associated food	No. of other foods with >10-fold lower levels	No. of other foods with <10-fold lower or higher levels	BFI concentration (after consumption of typical amount of food) is <10-fold higher than concentrations originating from confounders?	BFI concentrations (typical amount of food) are not >10-fold higher than endogenous concentrations?	Specificity (BFI Q)	Are there studies reporting the BFI occurrence/ content in the food source/ food?	Are there studies investigating possible non-food confounders for the BFI?	Specificity (DA)
11β,13-Dihydrolactucin	Lettuce	0	0	no	no	**1**	med	no	**C**
Arbutin	Pear	0	0	few cases	unk.	**1**	low	yes	**C**
N-Caproylhistamine	Tomato	0	0	no	no	**1**	high	yes	**A**
Salsolinol sulfate	Banana	0	1	few cases	unk.	**0.6**	med	yes	**B**
Phloretin*	Apple**	1	1	no	no	**0.6**	high	yes	**A**
Phloretin 2-*O*-glucuronide	Apple**	1	1	no	no	**0.6**	high	yes	**A**
Ergothioneine	Mushrooms	5	0	no	no	**0.6**	high	yes	**A**
Ergothioneine	Champignon	5	4	no	no	**0.4**	high	yes	**A**
Sulforaphane	Brassicaceae	0	0	no	no	**1**	high	yes	**A**
Sulforaphane-*N*-acetylcysteine	Brassicaceae	0	0	no	no	**1**	high	yes	**A**
Sulforaphane-*N*-acetylcysteine	Broccoli	2	5	no	no	**0.4**	high	yes	**A**
3,5-Dihydroxybenzoic acid	Whole Grain	0	0	no	no	**1**	high	yes	**A**
3,5-Dihydroxybenzoic acid	Rye	0	2	no	no	**0.6**	high	yes	**A**
3,5-Dihydroxybenzoic acid	Wheat	0	2	no	no	**0.6**	high	yes	**A**
Atractyligenin glucuronide	Coffee	0	0	no	unk.	**1**	med	no	**C**
Cyclo(L-leu-L-pro)	Coffee	0	4	unk.	no	**0.4**	low	yes	**C**
Caffeine	Coffee	0	10‡	no	no	**0**	high	yes	**A**
1-Methylxanthine	Coffee	0	3	no	unk.	**0.6**	high	yes	**A**
Trigonelline	Coffee	9	4	unk.	unk.	**0.13**	high	yes	**A**
Trigonelline	Green Beans	3	25	unk.	unk.	**0**	high	yes	**A**
Trigonelline	Peas	25	8	unk.	unk.	**0**	high	yes	**A**
Hesperetin	Orange	1	3	no	no	**0.6**	high	yes	**A**
Hesperetin 3-*O*-glucuronide	Orange	1	3	no	no	**0.6**	high	yes	**A**
Hesperetin 7-*O*-glucuronide	Orange	1	3	no	no	**0.6**	high	yes	**A**
Proline betaine	Orange	3	5	no	no	**0.4**	high	yes	**A**
Proline betaine	Citrus fruits	3	0	no	no	**0.6**	high	yes	**A**
Valencic acid glucuronide	Celery								

(*Continued*)

Table 1. Continued.

Part II (including the criteria: specificity, variability, dose-response).

		Variability							
BFI name	Associated food	Variability in food sources	Variability in foods	Is the BFI subject to a metabolism pathway?	Variability in humans	Variability (BFI Q)	Are there enough information to evaluate variability in plants and foods?	Are there studies that investigated intra- and/or inter-individual differences in BFI response?	Variability (DA)
11β,13-Dihydrolactucin	Lettuce	med	unk.	yes	unk.	**−0.17**	low	no	**C**
Arbutin	Pear	low	unk.	yes	unk.	**−0.17**	low	no	**C**
N-Caproylhistamine	Tomato	unk.	unk.	yes	low	**−0.17**	med	no	**C**
Salsolinol sulfate	Banana	med	unk.	yes	unk.	**−0.17**	low	no	**C**
Phloretin*	Apple	low	low	yes	med	**−0.33**	med	no	**C**
Phloretin 2-*O*-glucuronide	Apple	low	low	yes	unk.	**−0.17**	med	no	**C**
Ergothioneine	Mushrooms	low	low	yes	low	**−0.17**	high	no	**C**
Ergothioneine	Champignon	med	low	yes	low	**−0.17**	high	no	**C**
Sulforaphane	Brassicaceae	med	high	yes	low	**−0.5**	high	yes	**A**
Sulforaphane-*N*-acetylcysteine	Brassicaceae	med	high	yes	med	**−0.67**	high	yes	**A**
Sulforaphane-*N*-acetylcysteine	Broccoli	med	high	yes	unk.	**−0.5**	high	yes	**A**
3,5-Dihydroxybenzoic acid	Whole Grain	med	low	yes	med	**−0.33**	high	yes	**A**
3,5-Dihydroxybenzoic acid	Rye	low	low	yes	med	**−0.33**	high	yes	**A**
3,5-Dihydroxybenzoic acid	Wheat	low	low	yes	unk.	**−0.17**	high	yes	**A**
Atractyligenin glucuronide	Coffee	high	unk.	yes	unk.	**−0.5**	med	no	**C**
Cyclo(L-leu-L-pro)	Coffee	unk.	unk.	unk.	unk.	**0**	low	no	**D**
Caffeine	Coffee	med	med	yes	med	**−0.33**	high	no	**C**
1-Methylxanthine	Coffee	med	med	yes	unk.	**−0.17**	high	no	**C**
Trigonelline	Coffee	med	low	yes	unk.	**−0.17**	high	no	**C**
Trigonelline	Green Beans	unk.	unk.	yes	unk.	**−0.17**	low	no	**C**
Trigonelline	Peas	unk.	unk.	yes	unk.	**−0.17**	low	no	**C**
Hesperetin	Orange	low	low	yes	med	**−0.33**	low	no	**C**
Hesperetin 3-*O*-glucuronide	Orange	low	low	yes	med	**−0.33**	low	no	**C**
Hesperetin 7-*O*-glucuronide	Orange	low	low	yes	med	**−0.33**	low	yes	**C**
Proline betaine	Orange	unk.	unk.	unk.	med	**−0.17**	no data	yes	**D**
Proline betaine	Citrus fruits	unk.	unk.	unk.	med	**−0.17**	no data	yes	**D**
Valencic acid glucuronide	Celery								

		Dose-response							
BFI name	Associated food	No. of intervention studies with ≥3 doses demonstrating dose-response	No. of intervention studies with 1–2 doses demonstrating dose-response	Multi-study integration is possible that demonstrates a clear dose-response	No. of associations in observational studies with strong and concordant observational (r^2>0.5; p < 0.01 and SMD >0.5) evidence	No. of associations in observational studies with weak or moderate observational (r^2<0.5; p < 0.05 and SMD <0.5) evidence	$c_{baseline}$	Dose-response (BFI Q)	Dose-response (DA)
11β,13-Dihydrolactucin	Lettuce	0	2	no	0	0	zero	**0.2**	**C**
Arbutin	Pear	0	0	no	0	0	unk.	**0**	**D**
N-Caproylhistamine	Tomato	0	1	no	0	0	zero	**0.2**	**C**
Salsolinol sulfate	Banana	0	0	no	0	0	zero	**0.2**	**D**
Phloretin*	Apple	1	4	no	1	4	zero	**0.8**	**B**
Phloretin 2-*O*-glucuronide	Apple	0	3	no	0	0	zero	**0.2**	**C**
Ergothioneine	Mushrooms	0	0	no	0	3	unk.	**0.2**	**B**
Ergothioneine	Champignon	0	0	no	0	3	unk.	**0.2**	**B**
Sulforaphane	Brassicaceae	0	7	no	0	0	zero	**0.2**	**B**
Sulforaphane-*N*-acetylcysteine	Brassicaceae	0	10	no	0	0	zero	**0.2**	**B**
Sulforaphane-*N*-acetylcysteine	Broccoli	0	9	no	0	0	zero	**0.2**	**B**
3,5-Dihydroxybenzoic acid	Whole Grain	2	4	no	6	4	no	**0.8**	**A**
3,5-Dihydroxybenzoic acid	Rye	2	2	no	3	1	no	**0.8**	**A**
3,5-Dihydroxybenzoic acid	Wheat	0	1	no	0	2	no	**0.2**	**C**
Atractyligenin glucuronide	Coffee	0	0	no	0	2	unk.	**0.2**	**C**
Cyclo(L-leu-L-pro)	Coffee	0	0	no	0	2	unk.	**0.2**	**C**
Caffeine	Coffee	0	2	no	1	6	zero	**0.6**	**B**
1-Methylxanthine	Coffee	0	1	no	1	1	unk.	**0.4**	**C**
Trigonelline	Coffee	0	2	no	1	0	no	**0.4**	**C**
Trigonelline	Green Beans	0	0	no	0	0	unk.	**0**	**D**
Trigonelline	Peas	1	0	no	0	0	unk.	**0.6**	**C**
Hesperetin	Orange	1	5	no	0	1	zero	**0.8**	**B**
Hesperetin 3-*O*-glucuronide	Orange	0	0	no	0	0	unk.	**0**	**D**
Hesperetin 7-*O*-glucuronide	Orange	0	3	no	0	0	unk.	**0**	**C**
Proline betaine	Orange	2	3	no	2	4	low	**1**	**A**
Proline betaine	Citrus fruits	2	3	no	3	8	low	**1**	**A**
Valencic acid glucuronide	Celery								

(*Continued*)

Table 1. Continued.

Part III (including the criteria: robustness and analytical characterization).

		Robustness							
BFI name	Associated food	No. of human studies	No. of intervention studies (controlled diet)	No. of intervention studies (complex meal)	No. of observational studies (free living conditions)	Robustness (BFI Q)	Availability of observational studies or intervention studies with complex meals to evaluate the robustness of the BFI	Availability of intervention studies in which the BFI reflects food intake under controlled conditions	Robustness (DA)
11*β*,13-Dihydrolactucin	Lettuce	2	2	0	0	0.2	no	yes	C
Arbutin	Pear	1	1	0	0	0.2	no	yes	C
N-Caproylhistamine	Tomato	1	1	0	0	0.2	no	yes	C
Salsolinol sulfate	Banana	2	1	0	1	0.6	yes	yes	B
Phloretin*	Apple	8	5	0	3	1	yes	yes	A
Phloretin 2-*O*-glucuronide	Apple	7	6	0	1	1	yes	yes	B
Ergothioneine	Mushrooms	3	0	0	3	0.5	yes	no	C
Ergothioneine	Champignon	3	0	0	3	0.5	yes	no	C
Sulforaphane	Brassicaceae	7	7	0	0	0.4	no	yes	C
Sulforaphane-*N*-acetylcysteine	Brassicaceae	11	9	2	0	0.7	yes	yes	A
Sulforaphane-*N*-acetylcysteine	Broccoli	8	8	0	0	0.4	no	yes	C
3,5-Dihydroxybenzoic acid	Whole Grain	10	5	1	4	1	yes	yes	A
3,5-Dihydroxybenzoic acid	Rye	7	4	0	3	1	yes	yes	A
3,5-Dihydroxybenzoic acid	Wheat	3	1	0	2	0.8	yes	yes	B
Atractyligenin glucuronide	Coffee	3	1	0	2	0.8	yes	yes	B
Cyclo(L-leu-L-pro)	Coffee	2	0	0	2	0.4	yes	no	C
Caffeine	Coffee	9	1	1	7	1	yes	yes	B
1-Methylxanthine	Coffee	4	0	1	3	0.7	yes	yes	C
Trigonelline	Coffee	2	1	0	1	0.6	yes	yes	B
Trigonelline	Green Beans	2	1	0	1	0.6	yes	yes	B
Trigonelline	Peas	1	1	0	0	0.2	no	yes	C
Hesperetin	Orange	7	6	0	1	1	yes	yes	B
Hesperetin 3-*O*-glucuronide	Orange	3	2	0	1	0.8	yes	yes	B
Hesperetin 7-*O*-glucuronide	Orange	3	3	0	0	0.3	no	yes	C
Proline betaine	Orange	8	6	0	3	1	yes	yes	A
Proline betaine	Citrus fruits	11	7	0	5	1	yes	yes	A
Valencic acid glucuronide	Celery								

		Analytical Characterization											
BFI name	Associated food	Number of studies with qualitative methods	Number of studies with quantitative methods	Validation status	Short term stability	Long term stability	Analytical char. (BFI Q)	Are qualitative methods for BFI analysis available?	Are quantitative methods for BFI analysis available?	Was an authentic standard used?	Is the standard comer-cially available or is a synthesis described?	Are there validated methods available?	Are there studies about the BFIs stability in biological samples?	Analytical char. (DA)
11*β*,13-Dihydrolactucin	Lettuce	1	1	not validated	unk.	unk.	**0.2**	yes	yes	yes	comm.	yes	no	**B**
Arbutin	Pear	1	0	not validated	unk.	unk.	**0**	yes	no	no	comm.	yes	no	**D**
N-Caproylhistamine	Tomato	0	1	validated	unk.	unk.	**0.6**	no	yes	yes	syn.	yes	no	**B**
Salsolinol sulfate	Banana	2	0	not validated	unk.	unk.	**0**	yes	no	yes	syn.	yes	no	**B**
Phloretin*	Apple	0	8	validated	unk.	unk.	**0.6**	no	yes	yes	comm.	yes	no	**B**
Phloretin 2-*O*-glucuronide	Apple	4	3	validated	unk.	unk.	**0.6**	yes	yes	yes	comm.	yes	no	**B**
Ergothioneine	Mushrooms	2	0	not validated	unk.	unk.	**0**	yes	no	no	comm.	yes	no	**C**
Ergothioneine	Champignon	2	0	not validated	unk.	unk.	**0**	yes	no	no	comm.	yes	no	**C**
Sulforaphane	Brassicaceae	0	8	validated	stable	unk.	**0.7**	no	yes	yes	comm.	yes	yes	**A**
Sulforaphane-*N*-acetylcysteine	Brassicaceae	2	8	validated	stable	unk.	**0.7**	yes	yes	yes	comm.	yes	yes	**A**
Sulforaphane-*N*-acetylcysteine	Broccoli	2	8	validated	stable	unk.	**0.7**	yes	yes	yes	comm.	yes	yes	**A**
3,5-Dihydroxybenzoic acid	Whole Grain	0	9	validated	unk.	unk.	**0.6**	no	yes	yes	comm.	yes	no	**B**
3,5-Dihydroxybenzoic acid	Rye	0	9	validated	unk.	unk.	**0.6**	no	yes	yes	comm.	yes	no	**B**
3,5-Dihydroxybenzoic acid	Wheat	0	9	validated	unk.	unk.	**0.6**	no	yes	yes	comm.	yes	no	**B**
Atractyligenin glucuronide	Coffee	3	0	not validated	unk.	unk.	**0**	yes	no	yes	syn.	yes	no	**B**
Cyclo(L-leu-L-pro)	Coffee	2	0	not validated	unk.	unk.	**0**	yes	no	yes	comm.	yes	no	**B**
Caffeine	Coffee	7	2	validated	unk.	unk.	**0.6**	yes	yes	yes	comm.	yes	no	**B**
1-Methylxanthine	Coffee	4	0	not validated	unk.	unk.	**0**	yes	no	yes	comm.	yes	no	**B**
Trigonelline	Coffee	2	0	validated	unk.	unk.	**0.6**	yes	no	yes	comm.	yes	no	**B**
Trigonelline	Green Beans	1	1	validated	unk.	unk.	**0.6**	yes	yes	yes	comm.	yes	no	**B**
Trigonelline	Peas	1	0	validated	unk.	unk.	**0.6**	yes	no	yes	comm.	yes	no	**B**
Hesperetin	Orange	0	2	validated	unk.	unk.	**0.6**	no	yes	yes	comm.	yes	no	**B**
Hesperetin 3-*O*-glucuronide	Orange	1	0	not validated	unk.	unk.	**0**	yes	no	yes	comm.	yes	no	**B**
Hesperetin 7-*O*-glucuronide	Orange	0	3	validated	unk.	unk.	**0.6**	no	yes	yes	comm.	yes	no	**B**
Proline betaine	Orange	3	6	validated	unk.	unk.	**0.6**	yes	yes	yes	comm.	yes	no	**B**
Proline betaine	Citrus fruits	3	6	validated	unk.	unk.	**0.6**	yes	yes	yes	comm.	yes	no	**B**
Valencic acid glucuronide	Celery													

*Measured as aglycon after conjugate hydrolysis;
**including apple-derived foods;
‡caffeine containing soft drinks included; UC, unchanged; HM, host metabolite; GM, gut microbiota; HMGM, host-gut microbiota co-metabolite; unk., unknown; $c_{baseline}$, baseline concentration; com., commercially; syn., synthesis.

derivatives (Marks et al. Marks et al. 2009; Xiao et al. 2016). As the predominant metabolite, phloretin 2'-*O*-glucuronide is the most promising BFI candidate.

- An example for a BFI that is formed during food processing is the recently discovered cyclo(L-leu-L-pro) diketopiperazine, found correlated with coffee intake in observational studies (Guertin et al. 2015; Rothwell et al. 2019a; Shi et al. 2020). Cyclo(L-leu-L-pro) is not present as such in green coffee beans, but was reported to be formed through protein breakdown and cyclization of the linear dipeptide during the roasting process or during fermentation of raw coffee beans by microorganisms (Borthwick and Da Costa 2017; Rothwell et al. 2019b).

- An example for a BFI that failed the plausibility criterion is valencic acid (4-(3-methylbut-2-enoxy)benzoic acid) glucuronide. It was suggested as a putative BFI for celery intake in an intervention study (Lynn et al. 2019). However, no data were provided to confirm the presence of the precursor valencic acid in celery, and our literature survey (see methods section) did not identify any supporting data. Valencic acid glucuronide thus fails the plausibility criterion and should not be scored as a putative BFI without further investigation of its occurrence in celery.

Specificity criterion

The specificity of a BFI describes its ability to accurately reflect the consumption of a particular food or well-defined food group, without interference from other dietary sources or non-dietary factors. Specificity is included as a separate validation criterion to emphasize its importance. High specificity minimizes the risk of confounded dietary assessment in epidemiological studies, and therefore supports meaningful associations between health outcomes and dietary intake (Woodside et al. 2017; McNamara and Brennan 2020). To evaluate a BFIs specificity, it is essential to name the food it represents using its scientific name and indicate its processing level (e.g., raw, boiled, fermented). If a BFI reflects a food group, the foods included in the group must be clearly specified.

The first aspect to document is whether the BFI occurs in other foods, including those from the same species or taxonomic family. The BFI concentration, food consumption frequency and typical portion size must be compared for the target and the confounding foods. Regional variations in dietary habits, due to food availability or cultural influences may affect the generalizability of BFIs.

To investigate the food-related aspects, the literature survey should include studies performing an analytical screening of the BFI in various foods, ideally including different food groups, or comprehensive review articles. Results from at least two independent working groups should be compared, with careful examination of the methodology of original publications. To minimize the impact of extreme values, the range and median value are preferred over the mean. When a BFI is unlikely to occur in other foods due to a unique biosynthetic pathway, screening studies may be unnecessary, and the biosynthetic pathway itself can serve as evidence of specificity to a food or taxonomic family. Beyond its occurrence in other foods, the potential presence of the BFI in biological samples in non-food related contexts, such as medication, abuse of recreational drugs (e.g., alcohol, tobacco), and occupational or environmental chemical exposure should be examined in human studies. Databases such as the human metabolome database (HMDB; https://hmdb.ca/), EPA CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard), or DrugBank (https://go.drugbank.com) may be queried to find possible non-dietary origins of BFIs. Supplements, such as botanical extracts, may confound BFI specificity, but are not considered for specificity scoring due to their undefined composition and limited use at population level (IPSOS; European Public Affairs 2022). When applying a BFI in specific population contexts, potential implications of supplement intake should be considered case by case. Lastly, if BFIs are produced endogenously, their concentrations must be compared with the increase after a typical portion of the target food.

The calculation applied for the specificity sub-score (abbreviated with S in variables/formulas) is presented below. Figure 3 illustrates how to perform the data availability scoring (left side), and the BFI quality scoring (right side). A scheme for literature search on specificity is described in the Supplemental Material I.

Data availability (see left side of Figure 3)

Data availability is scored by considering the number and type of publications, the variety of foods covered in the studies, and the inclusion of research on non-dietary origins of the BFI. The availability of data on BFI occurrence in foods is classified as "high" (sub-score A) when more than two studies from independent groups are available, one or more being a large screening of foods that includes species other than the target food or a review summarizing the compound content in a range of foods. It is considered "medium" (maximum possible sub-score B) when at least two studies from independent groups are available, or when one study is a review or screening study. If there are no relevant studies or the studies come from only one group, or the screening covers only one plant species, the availability is rated "low" (maximum possible sub-score C). The maximum possible sub-score is also C if no data on non-food confounders are available. Relevant studies include observational studies explicitly addressing confounding factors in the association between food intake and BFI concentrations, or comprehensive literature reviews on the occurrence or formation of the BFIs across different research fields.

BFI quality (see right side of Figure 3)

The BFI quality sub-score is calculated based on the number of foods (including the target food) in which the BFI occurs and that are typically consumed in the target population: 5 points if the BFI is present in only the target food, 3 points if in 2 to 10 foods, and 1 point if it is present in more than 10 foods ($S_{occurrence}$). Foods with >100-fold lower BFI contents are not considered, even if consumed in considerable amounts, as they are unlikely to impair the BFI. Foods without quantitative data are not considered, as comparison with the BFI content in the target food is not possible. Foods not significantly consumed in the targeted population are not counted either. For European countries, food consumption data for a wide range of foods can be found in the EFSA Comprehensive European Food Consumption Database (https://www.efsa.europa.eu/en/data-report/food-consumption-data). Besides the number of foods the BFI occurs in, the sub-score is adjusted by a factor that depends on the relative BFI contents: a factor of 1.5 applies if all other foods have 10-fold lower levels compared to the target food, or if no more than three foods have similar levels (less than 10-fold lower), a factor of 1.0 applies if four to five foods have similar contents (less than 10-fold

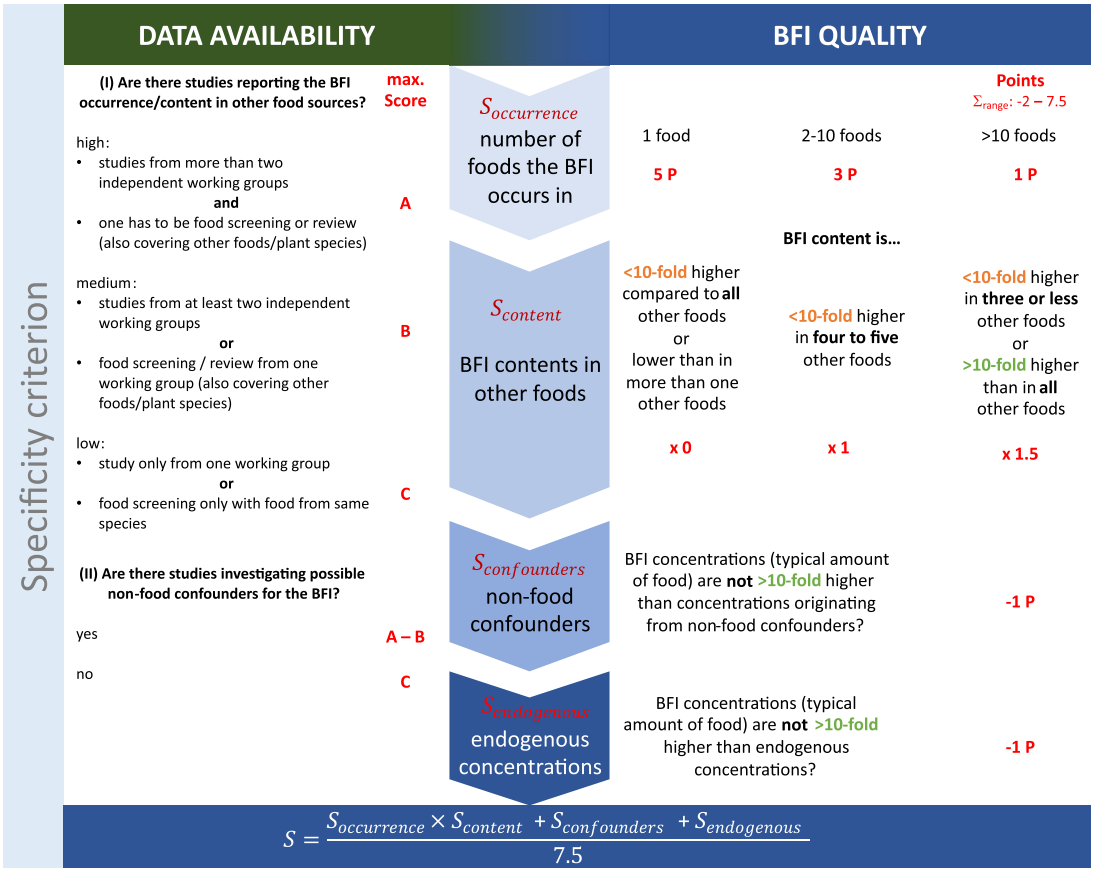

Figure 3. Detailed scheme of the specificity validation criterion. This validation criterion comprises of four aspects (center), which need to be considered for complete scoring of the BFI quality (right): (i) $S_{occurrence}$: number of foods the BFI occurs in, (ii) $S_{content}$: comparison of the BFI contents in other foods compared to the food in question, (iii) $S_{confounders}$: comparison of BFI concentrations in biological samples resulting from food intake and potential confounders, (iv) $S_{endogenous}$: comparison of BFI concentrations in biological samples resulting from food intake and potential endogenous concentrations. In red, the corresponding points (P) that can be awarded or subtracted for each aspect are displayed. The data availability (right) is evaluated based on (I) the number of studies reporting BFI contents in plants and foods, (II) and the availability of human studies that investigate possible confounders. The best achievable sub-score (red) is determined by the data availability level over all aspects (bold questions). The formula (bottom) shows the calculation of the relative number of points of the sub-score (S).

lower). The sub-score is 0 if more than five foods have similar content (less than 10-fold lower) or if any frequently consumed food has a higher concentration ($S_{content}$). When comparing BFI contents, ideally fresh weight (fw) is compared with fw and dry weight (dw) with dw. Otherwise, the typical water content of the food needs to be considered. Lastly, 1 point is subtracted if non-food confounders ($S_{confounders}$) or endogenous origin ($S_{endogenous}$) lead to concentrations in biological samples comparable (less than 10-fold lower; quantitative values required) to those observed after intake of a typical portion of the food. The absolute points ranging from −2 to 7.5 points are divided by maximum points (7.5) to obtain the relative points for the specificity sub-score.

In conclusion, evaluating the specificity of a BFI requires the unambiguous designation of the food and a detailed knowledge of the BFI's occurrence and contents in other foods, considering the influence of regional or population-specific dietary habits. Non-food related confounders and endogenous BFI concentrations are to be considered as well. The following examples illustrate key aspects for assessing BFI specificity, as well as potential pitfalls:

- The glycoside arbutin was discovered as a putative BFI for pear intake in a metabolomics study (Nieman et al. 2015). Its content in pears ranges from 0.1 to 1.0 mg/kg (fw), corresponding to 0.6 to 6.0 mg/kg (dw). Marjoram was shown to contain arbutin levels of 37 to 67 mg/kg (dw) (Lukas et al. 2010), but is excluded from the count due to its low frequency of consumption (0.2% for marjoram *vs*. 4.9% for pear; data from EFSA Comprehensive European Food Consumption Database (calculated mean for adults; 20.01.2025)) and its small portion size compared to pears (0.17 g for marjoram *vs*. 156 g for pear; same data source). Arbutin has been qualitatively detected but not properly quantified in apples, blueberries, walnuts, Chinese plums, and strawberry tree fruits (*Arbutus unedo*) (Eisele and Drake 2005; Saeedi et al. 2021; Morales 2022; Nahar et al. 2022). These food sources are not counted either due to the lack of quantitative data. Therefore, arbutin is relevant only for pears, awarding 5 points for the BFI quality sub-score. These 5 points are multiplied by 1.5 since arbutin is found in only one food, resulting in a relative specificity sub-score of 1.00, as no confounders or endogenous concentrations are reported to date. However, the data availability sub-score is C, as only qualitative screenings including other plant families

are available. A score of CI indicates that the BFI appears to be highly specific to pear based on current knowledge, but that its specificity has not yet been sufficiently investigated.

$$S = \frac{5 \times 1.5 + 0 + 0}{7.5} = 1.00$$

- The isothiocyanate sulforaphane (SFN) is a BFI for broccoli (single food) and/or Brassicaceae vegetables (food group) intake (Vermeulen et al. 2003). Sulforaphane is derived from the glucosinolate glucoraphanin, which is specific to the Brassicaceae family (Lamy et al. 2011). Glucoraphanin levels vary across Brassicaceae plants, with the highest contents in rocket (0.328 mg/g dw) and broccoli (0.140 mg/g dw), followed by cabbage (0.009 to 0.060 mg/g dw), brussels sprouts (0.031 mg/g dw), cauliflower (0.027 mg/g dw), kale (0.009 mg/g dw) and kohlrabi (0.004 mg/g dw). With the exception of kale and kohlrabi, all of the listed foods (n = 4) are considered for the specificity scoring, as the contents are higher or less than 10-fold lower compared to the content on broccoli. Due to this variation, sulforaphane is not highly specific to broccoli (sub-score: 0.40), but is highly specific for Brassicaceae vegetables as a food group (sub-score: 1.00).

$$S = \frac{3 \times 1 + 0 + 0}{7.5} = 0.40 \,|\, S = \frac{5 \times 1.5 + 0 + 0}{7.5} = 1.00$$

- Salsolinol sulfate, a BFI candidate for banana, is a phase II metabolite of salsolinol, itself formed through a reaction between dopamine present in banana and acetaldehyde during the ripening process (Riggin et al. 1976; Sojo et al. 2000; Yuan et al. 2017; Vazquez-Manjarrez et al. 2019). Another known source of salsolinol with similar content is cocoa, while other sources with lower contents are wine, beer and soy sauce (Riggin and Kissinger 1976; Duncan et al. 1984; Deng et al. 1997; Melzig et al. 2000). The conversion of dopamine into salsolinol can also occur *in vivo* in humans leading to endogenous concentrations. Salsolinol is controversially discussed as a biomarker for alcohol consumption in toxicology (Hipolito et al. 2012), with urine concentrations ranged from 0.1 to 13.3 ng/mL after ethanol intake (Musshoff et al. 1997). However, no quantitative data are currently available on salsolinol concentrations in biological samples after banana intake. Thus, no penalization is given for non-food confounders or endogenous concentration until quantitative data on salsolinol and its metabolites are obtained after banana intake. The specificity sub-score for the BFI quality is 0.60, and the data availability sub-score is B.

$$S = \frac{3 \times 1.5 + 0 + 0}{7.5} = 0.60$$

Variability criterion

BFI variability refers to the extent to which BFI concentrations vary in biological samples for the same level of food intake. The concept of biological variability was introduced for the first time in the validation scheme proposed by Cuparencu et al. (2024), while the variability of the BFI in the plant or food itself has never been considered. Both pre-consumption factors (e.g., variety, cultivation, storage, and processing) and individual-related factors (e.g., microbiota composition and functionality, genetics) can ultimately affect BFI concentrations in biological samples. Assessing the extent and sources of variability is crucial to evaluate the accuracy of food intake estimation from a given BFI and to identify biological variability that may limit its applicability (Brennan and Hu 2019; Wang et al. 2020; Abreu et al. 2021; Jang et al. 2021; Landberg et al. 2023). The initial source of variability originates from variability in the biomarker molecule in the food. Throughout food production and processing multiple factors can influence BFI content. In plants, variability may arise from variety, season, geography, agronomic conditions (e.g., soil composition, fertilization), plant health, ripeness, and the part consumed. Postharvest, storage conditions may further modify BFI levels, while processing steps such as washing, cutting, cooking, frying, roasting, or fermenting can affect them. Then, the meal composition and the food matrix can affect BFI bioavailability, as the same BFI may be absorbed to different extents depending on the food source.

After food ingestion, variability in BFI concentrations in biological samples can arise from intra-individual and inter-individual differences (Landberg et al. 2023). Factors influencing absorption, distribution, metabolism and excretion (ADME), such as the consumer's age, sex, genetics (e.g., single nucleotide polymorphisms for intestinal carriers or biotransformation enzymes), health status, physical activity, gut microbiota composition, as well as dietary habits are potential determinants for these differences (Jenab et al. 2009; Ottaviani et al. 2023; Parnell et al. 2025). Their relative contribution may vary across populations. The intraclass correlation coefficient (ICC) of repeated measurements over time quantifies the proportion of total variability attributable to intra-individual differences (Koo and Li 2016). It provides an estimate of how reliably a single biomarker measurement reflects an individual's average biomarker concentration (Landberg et al. 2023). Individual differences driven by genetics or gut microbiota composition can, in extreme cases, lead to strongly enhanced or absent metabolism, as exemplified by the metabolism of soy isoflavones, and ellagitannins (Morand et al. 2020). This phenomenon has given rise to the concept of "metabotyping", which categorizes metabolically similar individuals into subgroups, the so-called metabotypes (Riedl et al. 2017). Metabolized BFIs, especially those derived from microbial metabolism, tend to show higher inter-individual variability in biological samples than unmetabolized BFIs, and in the worst case, may lead to distinct metabotypes.

Figure 4 illustrates the scheme for the variability scoring. The relative points for the variability sub-score are abbreviated as V in the variables and formulas. The process begins by identifying the factors specifically relevant to the particular food and BFI. A scheme for search strings, which can be used to retrieve relevant publications, is described in the

Supplemental Material I. Next, the variability in the plant or food ($V_{food\ source}$) is assessed, followed by the evaluation of the variability in biological samples ($V_{metab.}$ and $V_{bio.sample}$).

Data availability (see left side of Figure 4)

To evaluate data availability, three questions must be examined and combined: food-related factors, BFI metabolism, and individual-related factors. When all food-related factors have been investigated in well-designed studies, data availability is considered "high", and the maximum possible sub-score is not restricted. If key factors remain unstudied or studies have poor design, the best achievable data availability sub-score is C, indicating "low" data availability.

Next, the data availability for metabolism is assessed. Suitable studies include, in decreasing order of interest, *in vivo* studies with food intake, *in vivo* studies with supplement intake, animal studies, and *in vitro* studies. A data availability sub-score of A is given when studies cover phase I, phase II, and gut microbiota metabolism. If one of these aspects is missing but relevant for the BFI, the best achievable sub-score is B or C, depending on the other factors. For evaluating intra- and inter-individual differences, both intervention and observational studies are suitable, although observational studies are less demonstrative due to the imprecision in dietary assessment. The best achievable sub-score is C when no studies have investigated intra- or inter-individual differences.

BFI quality (see right side of Figure 4)

First, the food-related variability is evaluated. In case the BFI is a metabolite, the variability is assessed for its precursor present in the plant or food. If one or more identified factors influencing variability can lead to a complete absence of the BFI in the food item, variability is considered "high" and subsequently the sub-score for variability in foods is penalized (−2 points). Provided a BFI is present across all varieties and forms of the food consumed, the BFI is considered acceptable and no penalty is imposed, even though variations in BFI contents may affect the BFIs ability to assess food intake. No points are deducted when the variability is unknown. Next, BFIs that are metabolized *in vivo* are penalized (−1 point). Intra-individual variability is considered "low" if the ICC is higher than 0.6 ("good" or "excellent" correlation), "medium" if between 0.4 and 0.6 ("fair" correlation), and "high" if below

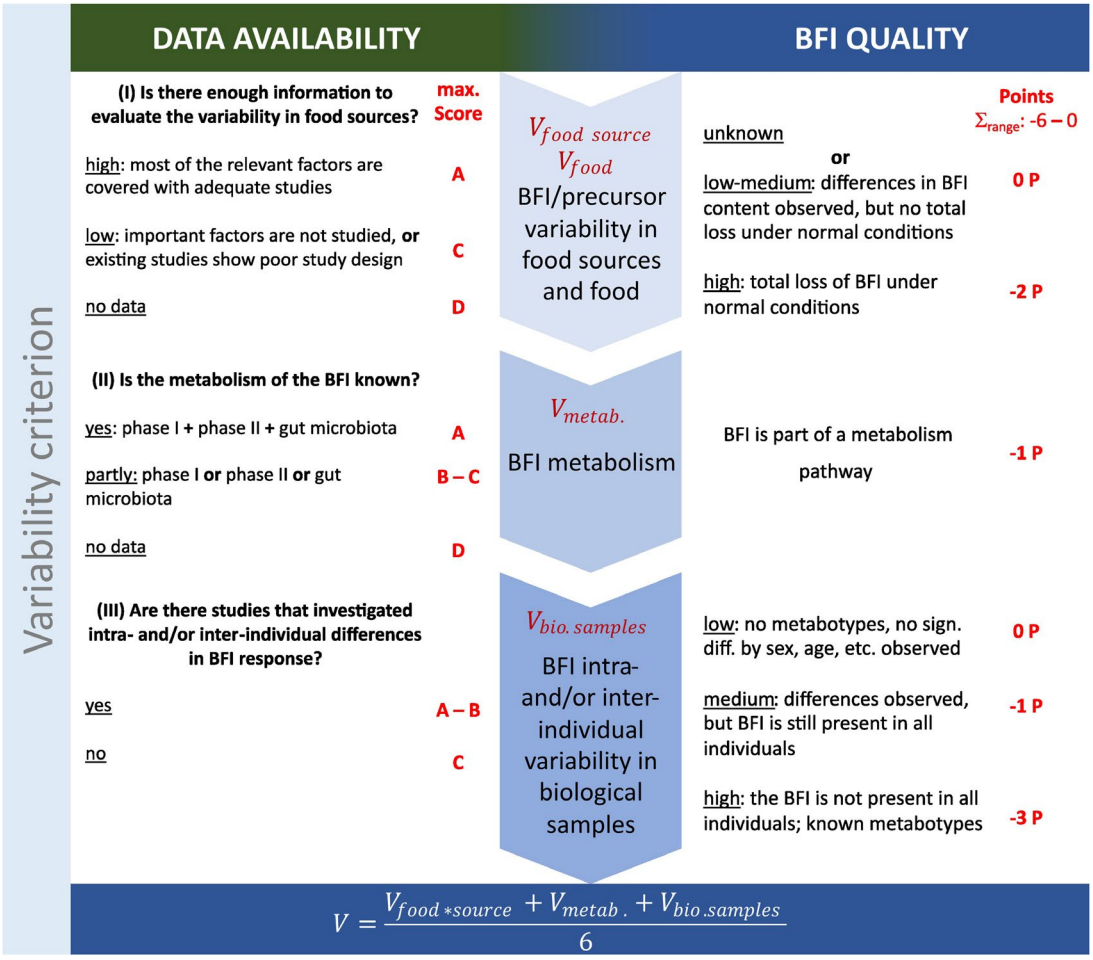

Figure 4. Detailed scheme of the variability validation criterion. This validation criterion comprises of three aspects (center) that need to be considered for complete BFI quality scoring (right): (i) $V_{food\ source/food}$: evaluation of the variability in BFI content in plants and foods, (ii) $V_{metab.}$: metabolism of the BFI, (iii) $V_{bio.samples}$: evaluation of the intra- and inter-individual variability in the BFI concentrations in biological samples. In red, the corresponding points (P) that can be subtracted for each aspect are displayed. The data availability (left) is evaluated based on (I) the degree of the investigation into factors affecting BFI content in plants and foods, (II) the knowledge about BFI metabolism (phase I, phase II, and gut microbiota), (III) and the availability of human studies that investigate possible factors causing intra- and/or inter-individual variability in BFI concentrations. The best achievable sub-score (red) is determined by the data availability level across all aspects (bold questions). The formula (bottom) shows the calculation of the relative number of points of the sub-score (V).

0.4, with significant differences between repeated measurements (Landberg et al. 2023). For inter-individual differences, variability is rated "low" if no metabotypes or major differences between subjects have been described (0 points subtracted), "medium" if differences exist, but the BFI is still present in all subjects (−1 point), and "high" if the BFI is absent in some individuals or if metabotypes are described (−3 points penalty). Also, no points are subtracted if variability is unknown. All penalization points are summed and the total, ranging from −6 to 0 points, is divided by 6 to obtain the relative points for the variability sub-score.

The following examples serve as a guide on how to assess the variability of a BFI and highlight possible pitfalls:

- Sulforaphane (SFN), found in cabbage, is formed from its precursor glucoraphanin by myrosinase. This reaction occurs when plant cells are damaged, such as by herbivore activity, processing, or chewing, allowing myrosinase to interact with glucoraphanin (Biondi et al. 2021; Costa-Perez et al. 2023). Cabbage is consumed in various forms (raw, cooked, fried, or fermented) and several factors, including washing, cutting, boiling (time and temperature), frying (time and temperature), and fermentation (time, bacteria, and ingredients), can affect the SFN content. Data availability scoring: Most of the factors that could potentially affect SFN content in the food have been thoroughly investigated. Also, the metabolism of SFN, including the human phase I, phase II metabolism, and the gut microbiota metabolism, is well understood. Furthermore, differences in the BFI response after food intake have been investigated in several studies. Therefore, the maximum sub-score is not restricted, resulting in a data availability sub-score A For BFI quality scoring, cooking method significantly affects SFN contents in cabbage and other Brassicaceae vegetables. During cooking, the heat-sensitive myrosinase is inactivated, and depending on cooking conditions, 75 to 100% of SFN can be lost (Tabart et al. 2018; Zayed et al. 2023). The variability in cooked cabbage is thus considered "high", resulting in a penalty (−2 points). As SFN can be generated by the gut microbiota from glucoraphanin that escapes hydrolysis by plant myrosinase (Angelino and Jeffery 2014; Oliviero et al. 2018), 1 additional point will be subtracted for V_{metab}. In human studies, substantial differences in C_{max} have been observed for similar food intakes. Genetic polymorphisms in genes coding for the human glutathione-S-transferase M1 (GSTM1) can influence the conjugation and the excretion of SFN (Gasper et al. 2005), yet without causing a complete loss of transferase activity. The variability is thus considered medium, and further penalized with −1 point. The final variability quality sub-score of SFN as a BFI for cooked cabbage intake is −0.67.

$$V = \frac{-2-1-1}{6} = -0.67$$

- 3,5-dihydroxybenzoic acid (DHBA) is an alkylresorcinol metabolite ($V_{metab.}$=-1) that is a well-studied BFI for whole grain intake (Jawhara et al. 2019; Landberg et al. 2019). The variability of the alkylresorcinol precursors in the food source is considered "medium" (varying contents between varieties) and in the food "low" ($V_{food\ source/food}$ = 0). Two studies reported ICC ranging from 0.32 to 0.55 in plasma and urine (Montonen et al. 2012; Wierzbicka et al. 2017), indicating an overall "fair" intra-individual variation for this BFI in both studies. Therefore, the intra-individual variability is rated as "medium". Regarding inter-individual differences, most studies did not find significant differences, except for one small intervention study with five participants that reported sex-dependent differences (Marklund et al. 2014). Thus, inter-individual variability is rated as "low". Considering both intra- and inter-individual differences, the overall rating follows the lowest rating, resulting in an overall rating of "medium" for $V_{bio.sample}$. As a result, 1 point is subtracted for the variability sub-score.

$$V = \frac{0-1-1}{6} = -0.33$$

- The metabolism of daidzein to equol represents a well-known example of metabotypes. Depending on gut microbiota composition, only some people can produce equol while others cannot, leading to high variability of equol concentration in biological samples for the same amount of soy-based food intake (Huser et al. 2018; Sri Harsha et al. 2018). As a result, $V_{bio.sample}$ is penalized with −3 points.

Dose-response criterion

A dose-response relationship describes the continuous increase of BFI concentrations with increasing food intake and is key for the application of the BFI in order to classify levels of consumption and, if possible, predict intake. From dose-response studies, calibration curves can be drawn to then be applied in observational studies, as demonstrated for proline betaine (Gibbons et al. 2017; D'Angelo et al. 2019; Gormley et al. 2022; Hu et al. 2024). Dose-response was already included in the FoodBAll validation scheme (Dragsted et al. 2018), but this is the first time that precise guidelines are provided to score it. The assessment of a BFI's dose-response requires consideration of both the study design and the strength of the statistical association.

Multi-dose intervention studies with well-characterized foods under controlled conditions are preferred for establishing a clear dose-response relationship. Dose-response studies should use realistic, everyday food amounts and include ≥3 clearly spaced doses. The difference between the next higher dose should be significant ($p \leq 0.01$) and ideally a calibration curve with a slope and coefficient of determination with $R^2 \geq 0.3$ is given. A plateau is acceptable, while U- or inverted U-shaped curves rule out a valid dose-response relationship. Furthermore, baseline concentrations (background),

minimum intake required to detect the BFI, and any saturation effects should be reported (Dragsted et al. 2018). Linear regression is the primary statistical approach, with non-linear regression applied when saturation occurs. Key indicators are the coefficient of determination (R^2) and the slope (β). Dose-response should be similar across independent studies.

In case only studies with one or two doses are available, a multi-study integration approach can be used to evaluate the dose-response relationship, provided that (i) quantitative data is available and comparable (possible restrictions: different normalization of urine), (ii) comparable sampling time or kinetic is available, (iii) ingested food is comparable (possible bias: different varieties), and (iv) potential interference with other dietary components dependent on study design is minimal. To date, no multi-study integration approach has been applied, however, this is considered a potential future option and therefore included.

Observational studies can also be used to evaluate dose-response. Typically, BFI concentrations in biological samples are related to continuous intake data (FFQ, dietary records (DR), or 24-hr recalls) using Pearson or Spearman correlations and linear regression. The strength of the relationship is assessed using the correlation coefficient (ρ or r; $r > 0.5$ "good", $r = 0.2–0.5$ "fair", $r < 0.2$ "poor"), the explained variance (R^2), and the significance of the slope (p-value) (Schober et al. 2018). When intake is categorized (tertiles, quartiles, quintiles), biomarker levels can be compared using ANOVA (or Kruskal-Wallis for non-parametric data), combined with a linear trend test or a regression on mean or median intake values of each category. Significance is assessed as p-trend (<0.05), while effect size is characterized by the standardized mean difference (SMD > 0.5) or the Q4/Q1 ratio (≥1.3) (Gallardo-Gómez et al. 2024). Dose-response established in observational data should be interpreted with caution, as questionnaire data are imprecise due to self-reporting bias (Hedrick et al. 2012; Cui et al. 2021; Ottaviani et al. 2024).

Another important aspect is the baseline concentration, which is expected to be negligible for a good BFI but might vary across studies depending on the study design, or the applied methodology. The BFI should not be detectable ($c_{baseline} = 0$) or at least >10-fold lower than concentrations normally observed after intake of the target food or other sources. In the case of differing results (e.g., $c_{baseline} = 0$ *vs.* $c_{baseline} \neq 0$), the discrepancy must be explainable, e.g., the quantification using methods with varying sensitivity can lead to either detection or non-detection of a compound's concentration in a sample (Dragsted et al. 2018).

Figure 5 illustrates the different steps to score the dose-response criterion (relative points for the sub-score are abbreviated with DR in variables/formulas).

Data availability (see left side Figure 5)

The data availability is determined by (i) the number of studies (observational studies or intervention studies) available to evaluate the dose-response, and (ii) the number of doses in intervention studies, with 4 levels from A to D.

BFI quality (see right side Figure 5)

When two or more intervention studies with three or more doses show a strong monotonic dose-response relationship,

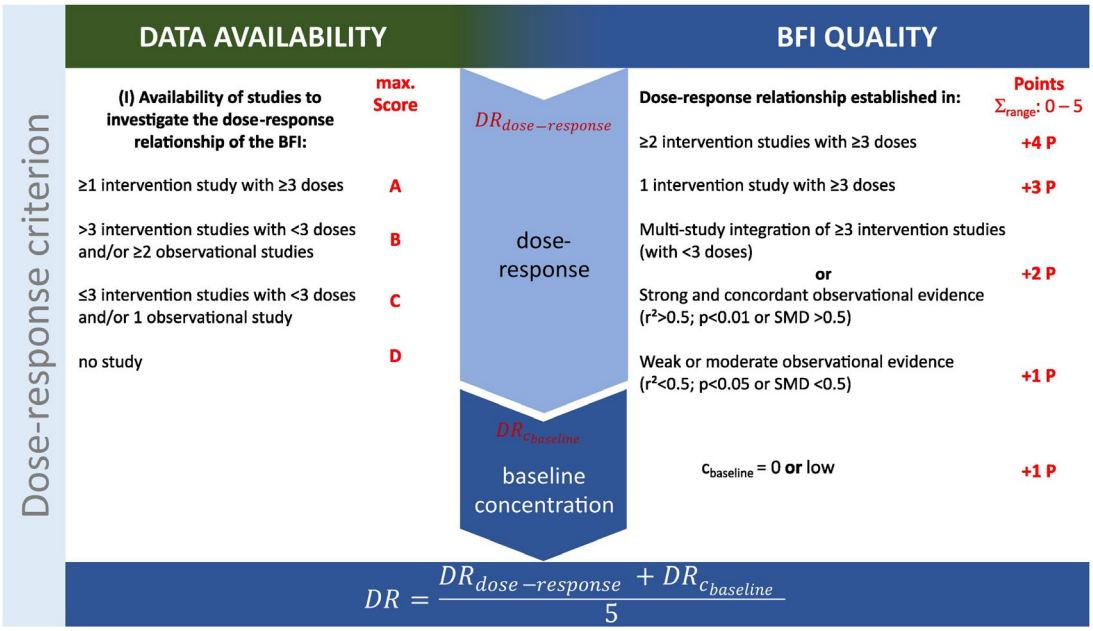

Figure 5. Detailed scheme of the dose-response validation criterion, showing all possible results (marked in red). This validation criterion comprises of two aspects (center) that need to be considered for complete scoring of the BFI quality (right): (i) $DR_{dose\text{-}resp.}$: the BFI shows a clear dose-response – the BFI concentration in biological samples rises with increasing amounts of food intake, (ii) $DR_{cbaseline}$: the BFI baseline concentration is zero or very low compared to the concentrations reached after intake of the target food. The data availability (left) is evaluated based on the number of intervention studies that involved a sufficient number of dose levels, or are comparable enough to allow a multi-study integration, and on the number of observational studies that applied appropriate statistical approaches (e.g., Pearson or Spearman correlation analyses, linear regression, p-trend test, or standardized mean difference (SMD, Q1 *vs.* Q4)). The formula (bottom) shows the calculation of the relative number of points of the sub-score (DR).

4 points are awarded ($DR_{dose\text{-}response}$), versus only 3 points when just one such study is available. Two points are awarded when a strong dose-response relationship was established in at least one observational study providing strong relationship ($r > 0.5$; $p < 0.01$; SMD > 0.5) or by integration of three or more intervention studies with less than three doses. In case an observational study provides weak or moderate evidence ($r < 0.5$; $p < 0.05$; SMD < 0.5), 1 point is given. Furthermore, when baseline concentrations in intervention studies are zero or low, or non-consumers in observational studies do not have detectable BFI concentrations, 1 point is awarded ($DR_{cbaseline}$). The sum of assigned points (0–5) is divided by maximum possible sub-score to obtain the relative dose-response sub-score.

The following examples illustrate how to apply the scoring system:

- An extensively studied example is proline betaine, a biomarker for orange intake. For scoring the data availability for dose-response, four observational studies (r = 0.15 to 0.54) (Wang et al. 2018; 2020; French et al. 2023; Hu et al. 2024) and five intervention studies (Atkinson et al. 2007; Heinzmann et al. 2010; Gibbons et al. 2017; Lang et al. 2017; Saenger et al. 2021), of which two have 3 doses (Gibbons et al. 2017; Saenger et al. 2021), and three have 1 dose are considered suitable (Atkinson et al. 2007; Heinzmann et al. 2010; Lang et al. 2017). The sub-score is A. The BFI quality $DR_{dose\text{-}resp.}$ was scored 4, as a clear dose-response relationship was demonstrated independently in two intervention studies (Gibbons et al. 2017; Saenger et al. 2021) and one observational study (Lloyd et al. 2011). The baseline concentration in intervention studies ranged from 0 to 8% compared to the maximum concentration reached after orange intake (Atkinson et al. 2007; Lang et al. 2017; Saenger et al. 2021), indicating that baseline levels may only interfere for low levels of intake ($DR_{cbaseline}$ = 1). The sub-score for BFI quality is 1.00.
- An example of a recently discovered BFI is the lettuce-specific biomarker 11β,13-dihydrolactucin. Besides an intervention study with an untargeted metabolomics approach (García et al. 2020), only one intervention study with quantitative data is available (Weng et al. 2020). Both studies only used one dose. No observational studies are available, resulting in a data availability sub-score of C. Adequate studies to assess the BFI quality for dose-response are lacking ($DR_{dose\text{-}resp.}$ = 0). The baseline concentration was reported as not detectable ($DR_{cbaseline}$ = 1). Finally, the BFI quality sub-score is 0.20.

Robustness criterion

Robustness is a key validation criteria already included in previous frameworks. A robust BFI consistently performs well under various conditions without significant deterioration in quality or performance, i.e., reflecting the intake of the target food in a free-living population, despite various dietary backgrounds and external factors that may influence the BFI (Dragsted et al. 2018). To achieve a high level of robustness, BFIs should be confirmed in both controlled dietary intervention studies and, importantly, in observational studies (Dragsted et al. 2018). Moreover, the selected studies should meet common reporting standards (Ulaszewska et al. 2019; Rigutto-Farebrother et al. 2023).

With respect to robustness, intervention studies are important for studying the BFI response under controlled conditions with a known and well-characterized food, complex meals or habitual diet. Observational studies demonstrate the robustness of BFIs in free-living conditions. They can reveal if a BFI is potentially influenced by other foods or non-food confounders such as smoking, alcohol consumption, or medication. The BFI should either not be detected or be present at very low concentrations in individuals who did not consume the target food. Even if much lower concentrated than the target food, some interfering foods may still elevate baseline BFI concentrations in biological samples.

Robustness is linked to variability: while high variability inevitably reduces robustness, low variability does not guarantee high robustness, as robustness also depends on additional factors. Figure 6 illustrates the different steps to score the BFI robustness (relative points for the sub-score are abbreviated with R in variables/formulas).

Data availability (see left side Figure 6)

It is determined by (i) the number of intervention studies showing the BFI response following the intake of the food in a fully controlled setting, and (ii) the number of studies (observational studies or intervention studies with complex meals) available to assess robustness in the context of a complex diet. The highest sub-score A is given when more than one controlled intervention study and more than one observational study are available. If only one observational study or one intervention study with complex meals is available, regardless of the number (equal or greater than one) of intervention studies, the sub-score is B, or *vice versa* with respect to intervention studies. If only one study type (controlled intervention or observational study) is available, the sub-score is limited to C.

BFI quality (see right side Figure 6)

First, all human studies (intervention and observational studies) of sufficient quality in which the BFI reflects the food intake are identified. Points are awarded when: (i) the BFI is shown to reflect food intake in a highly controlled diet (single food) (+1 point), (ii) the BFI reflects food intake in complex meals or under free-living conditions (+3 points) (R_{meal}). Suitable observational studies typically report a "fair" or "good" correlation between BFI concentrations and food intake as measured by FFQ, DR or 24-hr recalls. Other statistical approaches, such as ROC curves may be used. Additional points (1–3) are awarded depending on the total number of relevant studies ($R_{studies}$). The points are multiplied by 2 if both intervention and observational studies are available ($R_{int.+obs.}$). The sum of assigned points (1 to 10)

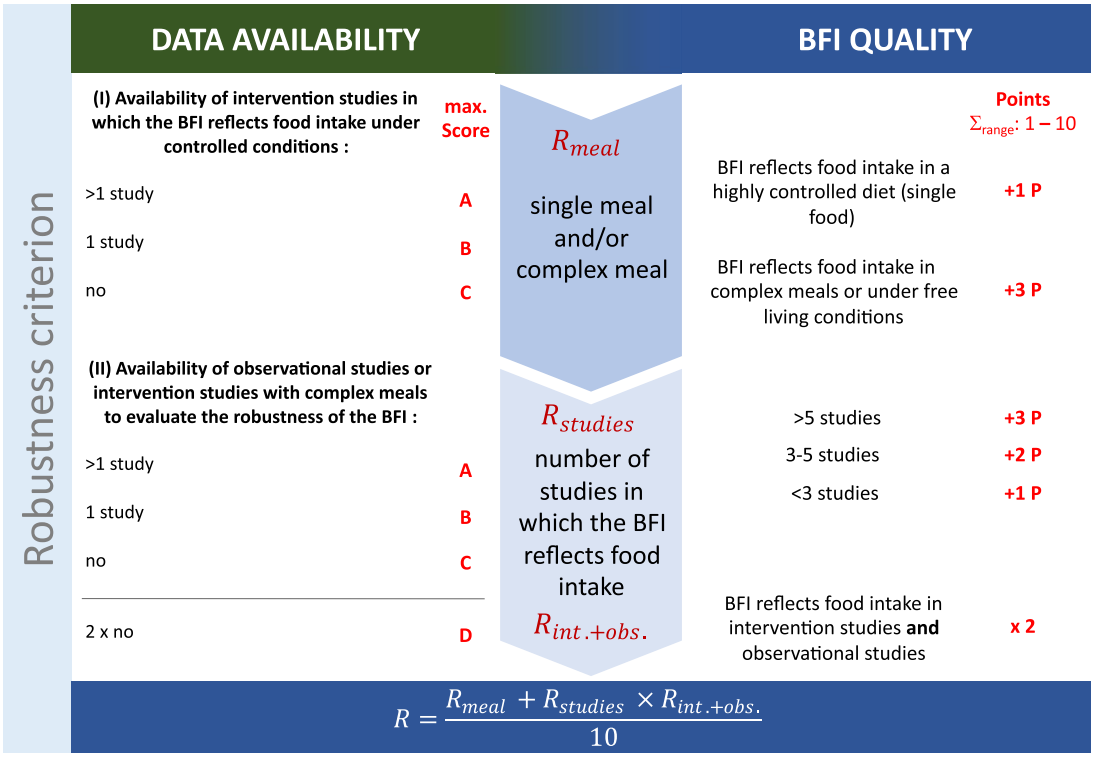

Figure 6. Detailed scheme of the robustness validation criterion. This validation criterion comprises of two aspects (center) that need to be considered for complete scoring of the BFI quality (right): (i) R_{meal}: the BFI has been shown to reflect the intake of the target food under highly controlled dietary conditions and/or in complex meals or under free living conditions. (ii) $R_{studies}$: number of studies in which the BFI reflects the intake of the target food. The data availability (left) is evaluated based on (I) the number of intervention studies in which the BFI reflects food intake under controlled conditions, and (II) the number of observational studies or intervention studies with complex meals. The best achievable sub-score (red) is determined by the data availability level over all aspects (bold questions). The formula (bottom) shows the calculation of the relative number of points of the sub-score (R).

is divided by the maximum possible sub-score to obtain the relative points for the robustness sub-score.

The following examples illustrate how to apply the scoring system:

- Proline betaine is an extensively studied biomarker for orange intake. Four observational studies (Pujos-Guillot et al. 2013; Wang et al. 2018; 2020; Hu et al. 2024) and six intervention studies (Atkinson et al. 2007; Heinzmann et al. 2010; Pujos-Guillot et al. 2013; Gibbons et al. 2017; Lang et al. 2017; Saenger et al. 2021) showed the proline betaine-orange intake relationship under controlled and free-living conditions. The data availability sub-score is A. Scoring the BFI quality: All intervention studies were single food interventions and demonstrated the BFI origin under highly controlled conditions. As a result, 1 point is awarded. So far, no intervention study with a complex meal is available. However, as "fair to good" correlations were observed under free living conditions in four observational studies (Pujos-Guillot et al. 2013; Wang et al. 2018; 2020; Hu et al. 2024), 3 additional points are awarded (R_{meal} = 4). As more than five studies demonstrated a robust association between orange intake and proline betaine (Atkinson et al. 2007; Heinzmann et al. 2010; Pujos-Guillot et al. 2013; Gibbons et al. 2017; Lang et al. 2017; Wang et al. 2018; 2020; Saenger et al. 2021; Hu et al. 2024), 3 points ($R_{studies}$ = 3) are awarded. These points are multiplied by 2 ($R_{int.+obs}$ = 2), since both interventional and observational studies exist, resulting in a total of 6 points. The sub-score for BFI quality is 1.00.

$$R = \frac{(1+3)+3\times 2}{10} = 1.00$$

- An example of a recently discovered BFI is the tomato-specific biomarker *N*-caproylhistamine. Following initial discovery in a preliminary intervention study, its dose- and time-responses were examined in a second controlled intervention study (Hövelmann et al. 2020). However, to date *N*-caproylhistamine has not been confirmed in any observational study or intervention study with complex meals, leaving open the question of whether it performs well in a free-living population. According to the scoring system, the robustness quality sub-score is 0.20, and the data availability sub-score is C.

$$R = \frac{(1+0)+1\times 1}{10} = 0.20$$

Analytical characterization criterion

Ideally, a BFI should be stable and quantitatively measured with high accuracy and precision in biological samples. The

analytical characterization of a BFI refers to the validation of the analytical method(s) used to measure the BFI in biological samples and the specific characteristics of the BFI itself. Analytical characterization encompasses the three analytical criteria considered in the FoodBAll validation scheme (Dragsted et al. 2018).

Many BFIs have been discovered in untargeted metabolomics studies. However, a targeted method of analysis, typically with a calibration curve using commercially available authentic standards, is required for their quantification. According to the guidelines from the Food and Drug Administration (FDA) of the U.S. Department of Health and Human Services (U.S. Department of Health and Human Services Food and Drug Administration 2018) or the International Council for Harmonisation (ICH, 2022), validation parameters to assess quantification methods are the specificity/selectivity, linearity, accuracy, precision (with repeatability and reproducibility), percentage of recovery, absence of matrix effects, limit of detection (LOD) and limit of quantification (LOQ) (Kruve et al. 2015a, 2015b; Raposo and Ibelli-Bianco 2020). Sample preparation protocols (e.g., extraction, purification, enrichment, derivatization, or enzymatic/chemical sample treatment) may significantly influence the method performance and reproducibility and should be carefully considered. For multi-target methods comprising a larger number of analytes, no guidelines have been proposed so far. Ideally, a method should be validated across different laboratories to ensure consistent and equivalent measurements of the BFI (inter-lab validation). Another relevant aspect is the stability of the BFI in biological samples over both short- and long-term periods. Short-term stability refers to the time from sampling to sample preparation and measurement with only a short duration of storage, typically a few days. The BFI could be affected, for example, by oxidation, light, enzymes, or labware itself (e.g., adhesion of compounds to plastic or glass). Long-term stability refers to the stability of the BFI over extended periods (months or years) during storage of the biological samples. Parameters that influence long-term stability include storage temperature (−20 °C, −80 °C or liquid nitrogen), storage time, or freeze and thaw cycles (Committee for Medicinal Products for Human Use and European Medicines Agency 2011; U.S. Department of Health and Human Services Food and Drug Administration 2018). Assessing long-term stability is essential for reliable BFI measurement in stored samples from longitudinal or biobank-based studies.

Figure 7 illustrates the scoring of analytical characterization, with the relative number of points for the analytical characterization sub-score abbreviated as AC.

Data availability (see left side of Figure 7)

A sub-score of D is given when only one qualitative method is available. In case at least two independent studies describe non-validated quantitative methods or the studies do not meet the reporting standards for method validation (Kruve et al. 2015a, 2015b; Raposo and Ibelli-Bianco 2020), the maximum attainable sub-score is C, even if a reference standard is used. When information about the method validation or the BFI stability was investigated, the sub-score is B. If all aspects have been covered the sub-score is A.

BFI quality (see right side of Figure 7)

The methods employed to measure the BFI in the human studies previously listed for scoring its robustness, can be considered and complemented by other analytical articles to assess the BFI analytical characterization (note that method development and validation can be published separately): 4 points are awarded if the validated method (with all of the above-mentioned parameters investigated in the validation process) has also undergone successful inter-lab validation, 3 points if the method is validated in a single laboratory, and 1 point for quantitative methods that use reference standards and calibration curves, but for which the validation of all above-mentioned parameters is not described. No points are awarded for qualitative methods (AC_{method}). Additional points are assigned when the BFI is shown to be stable over a short-term (+0.5 point) and/or over a long-term (+0.5 point) scale. If stability is unknown no points are awarded, and if the BFI is known to be unstable the sub-score is penalized (−1 point for each short- and long-term stability; $AC_{stability}$). If a BFI comes from an unstable compound class but lacks BFI-specific studies, literature on similar compounds can be considered to evaluate the expected stability. The absolute points (−2 to 5) points are divided by maximum possible sub-score to obtain the relative points for the analytical characterization sub-score.

The following examples are given to illustrate BFIs with different analytical characterization sub-scores:

- *N*-caproylhistamine, a BFI for tomato intake, was investigated in an intervention study using a validated method and an authentic reference standard synthesized by the same research group. All validation criteria were fulfilled for the method (AC_{method} = 3), but the stability of *N*-caproylhistamine was not investigated ($AC_{stability}$ = 0) (Hövelmann et al. 2020). The BFI quality sub-score is thus 0.60 and the data availability is B.

$$AC = \frac{3 + (0 + 0)}{5} = 0.60$$

- Sulforaphane-*N*-acetyl cysteine (SFN-NAC) was analyzed in numerous studies with a range of qualitative (Andersen et al. 2013; 2014) and quantitative methods (Dominguez-Perles et al. 2014; Sun et al. 2020; Baenas et al. 2017; Al Janobi et al. 2006; Charron et al. 2018; Vermeulen et al. 2006; J.D. Clarke et al. 2011; Conaway et al. 2000; Bouranis et al. 2021), mostly using authentic reference standards. Three of the quantitative methods are also validated according to the above-mentioned validation guidelines (Al Janobi et al. 2006; Vermeulen et al. 2006; Dominguez-Perles et al. 2014). Al Janobi et al. (2006) investigated the influence of pH value and temperature on short-term stability of SFN metabolites in

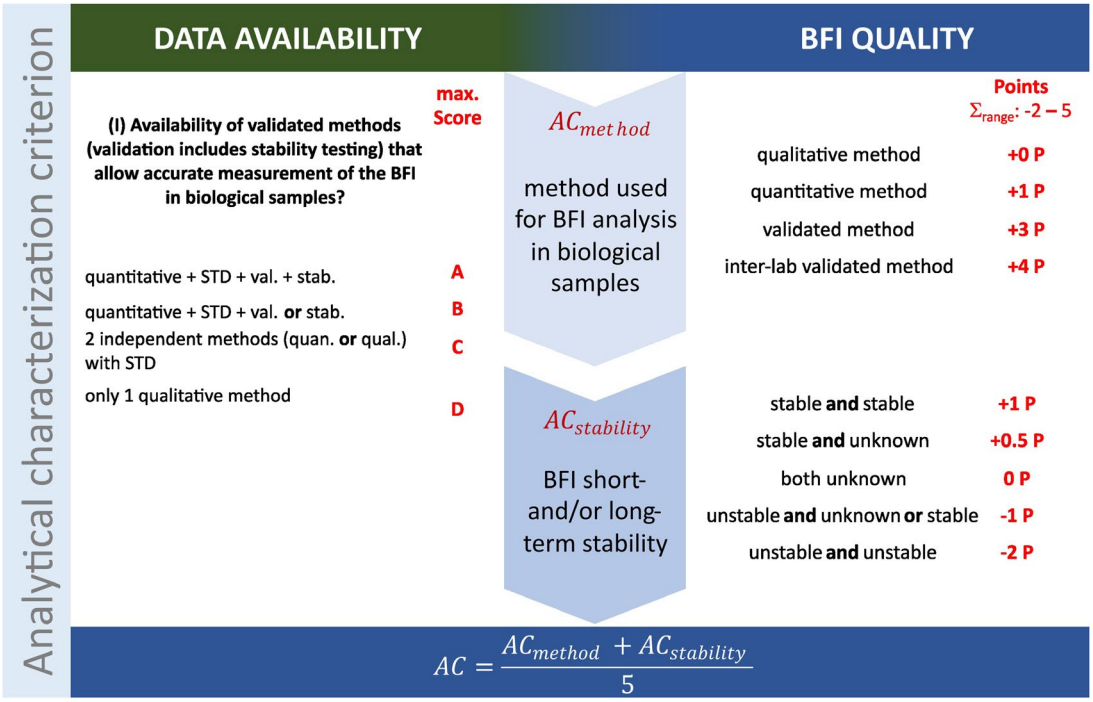

Figure 7. Detailed scheme of the analytical characterization validation criterion. This validation criterion comprises of two aspects that need to be considered to complete scoring of the BFI quality (right): (i) AC_{method}: evaluation of the validation status of the applied methods used in human studies to detect/quantify the BFI. Qualitative methods only provide relative values, whereas quantitative methods provide absolute concentration of the BFI in biological samples. To ensure precise and reproducible measurements, methods are traditionally validated including parameters such as: selectivity, linearity, accuracy, precision, recovery, matrix effects, and limit of quantification. When the validation parameters are consistent across different labs, the methods are inter-lab validated. The evaluation refers to the method with the highest validation status across all methods available, (ii) $AC_{stability}$: evaluation of the stability of the BFI on a short- and long-time scale. The data availability (left) is evaluated based on the quality, validation status, and number of methods applied in human studies. The best achievable sub-score (red) is determined by the data availability level over all aspects (bold questions). The formula (bottom) shows the calculation of the relative number of points of the sub-score (AC).

biological samples. SFN-NAC was stable at 4 °C at a pH range from 2 to 7.4. The long-term stability has not been investigated in any study so far. Therefore, 0.5 points are awarded for $AC_{stability}$. This results in a sub-score of AI (0.70).

$$AP = \frac{3 + (0.5 + 0)}{5} = \frac{3.5}{5} = 0.70$$

Final score calculation

The global data availability score of a BFI is calculated from all data availability sub-scores using the weighted ordinal-median method, which gives a central (median) category rather than the most frequent one. It uses the ordered data availability levels A > B > C > D and their respective weights. The weights assigned for each level are summed, then the cumulative totals calculated from A (highest level) to D (lowest level). The weighted median is the first level whose cumulative weight equals or exceeds half of the total weight (½ × 13 = 6.5). For example, for phloretin 2-*O*-glucuronide (S:A/V:C/DR:C/R:B/AC:B; weight 3, 3, 3, 3, 1), the weight sums are: A = 3, B = 4, C = 6, D = 0; cumulative weights are: A = 3, B = 7, C = 13. As the first cumulative weight exceeding 6.5 is B, the overall data availability score is B.

The global BFI quality score is calculated by multiplying the numerical sub-scores specificity (S), variability (V), dose-response (DR), robustness (R), and analytical characterization (AC) by the corresponding weighting factors and dividing by the maximum reachable points for each sub-score.

$$Overall\ BFI\ Quality = \frac{3 \times S + 3 \times V + 3 \times DR + 3 \times R + 1 \times AC}{13}$$

The resulting score is rounded to two decimals and then categorized into four levels. The quality score ranges from −0.34 to 0.77: Level I ≥ 0.57, Level II 0.56-0.33, Level III 0.32-0.20, Level IV <0.20 (see methods section for the definition of thresholds). The resulting levels provide an easy to interpret categorization of BFI quality. Full numeric scores are reported in Table 2 and can also be used to distinguish BFIs within each level.

To keep the information on individual validation criteria visible and presented in a standardized, expert-interpretable format, all sub-scores for BFI quality and data availability (*) are combined into a single string without spaces following this structure:

overall score(specificity:S*S/variability:V*V/dose-response: DR*DR/robustness:R*R/analytical characterization:AC*AC).

Example for phloretin 2-*O*-glucuronide: BII(S:AII/V:CI/DR:CIV/R:BI/AC:BII)

The string offers detailed insight into each sub-score, thereby facilitating the identification of potential limitations and knowledge gaps related to a given BFI.

Table 2. Summary showing the common name of the BFI, associated food, the result of pass-or-fail criteria, sub-scores for the five scoring criteria (specificity (S), variability (V), dose-response (DR), robustness (R), and analytical characterization (AC)), the overall score, and the experts' view. For the sub-scores the data availability (DA) is shown in the first column, followed by the BFI quality score and relative number of points for the BFI quality. For the overall score, first the data availability score is reported, then, the calculated relative BFI quality score with the corresponding level followed by the expert view and the string (see section: Final score calculation). The overall score for DA, BFI quality, the expert view and string are in bold.

			Sub-scores																Overall score					
BFI name	Associated food	Fail-or-pass	Specificity			Variability			Dose-response			Robustness			Analytical char.			DA	BFI quality		Expert view		String
11β,13-Dihydrolactucin	Lettuce	Pass	C	I	1.00	C	I	−0.17	C	IV	0.20	C	IV	0.20	B	III	0.20	**C**	0.30	**III**	**C**	**II**	**CIII(S:CI/V:CI/DR:CIV/R:CIV/AC:BIII)**
Arbutin	Pear	Pass	C	I	1.00	C	I	−0.17	D	IV	0.00	C	IV	0.20	D	IV	0.00	**C**	0.24	**III**	**B-C**	**II**	**CIII(S:CI/V:CI/DR:DIV/R:CIV/AC:DIV)**
N-Caproylhistamine	Tomato	Pass	A	I	1.00	C	I	−0.17	C	IV	0.20	C	IV	0.20	B	II	0.60	**C**	0.33	**II**	**B**	**II**	**CII(S:AI/V:CI/DR:CIV/R:CIV/AC:BII)**
Salsolinol sulfate	Banana	Pass	B	II	0.60	C	I	−0.17	D	IV	0.20	B	II	0.60	B	IV	0.00	**B**	0.28	**III**	**B**	**II**	**BIII(S:BII/V:CI/DR:DIV/R:BII/AC:BIV)**
Phloretin*	Apple**	Pass	A	II	0.60	C	II	−0.33	B	I	0.80	A	I	1.00	B	II	0.60	**B**	0.52	**II**	**A**	**I**	**BII(S:AII/V:CII/DR:BI/R:AI/AC:BII)**
Phloretin 2-*O*-glucuronide	Apple**	Pass	A	II	0.60	C	I	−0.17	C	IV	0.20	B	I	1.00	B	II	0.60	**B**	0.42	**II**	**A**	**I**	**BII(S:AII/V:CI/DR:CIV/R:BI/AC:BII)**
Ergothioneine	Mushrooms	Pass	A	II	0.60	C	I	−0.17	B	IV	0.20	C	II	0.50	C	IV	0.00	**C**	0.26	**III**	**C**	**II**	**CIII(S:AII/V:CI/DR:BIV/R:CII/AC:CIV)**
Ergothioneine	Champignon	Pass	A	II	0.40	C	I	−0.17	B	IV	0.20	C	II	0.50	C	IV	0.00	**C**	0.22	**III**	**C**	**II**	**CIII(S:AII/V:CI/DR:BIV/R:CII/AC:CIV)**
Sulforaphane	Brassicaceae	Pass	A	I	1.00	A	II	−0.50	B	IV	0.20	C	III	0.40	A	I	0.70	**A**	0.31	**III**	**A**	**I**	**AIII(S:AI/V:AII/DR:BIV/R:CIII/AC:AI)**
Sulforaphane-*N*-acetylcysteine	Brassicaceae	Pass	A	I	1.00	A	III	−0.67	B	IV	0.20	A	II	0.70	A	I	0.70	**A**	0.42	**II**	**A**	**I-II**	**AII(S:AI/V:AIII/DR:BIV/R:AII/AC:AI)**
Sulforaphane-*N*-acetylcysteine	Broccoli	Pass	A	II	0.40	A	II	−0.50	B	IV	0.20	C	III	0.40	A	I	0.70	**A**	0.25	**III**	**A**	**I**	**AIII(S:AII/V:AII/DR:BIV/R:CIII/AC:AI)**
3,5-Dihydroxybenzoic acid	Whole grain	Pass	A	I	1.00	A	II	−0.33	A	I	0.80	A	I	1.00	B	II	0.60	**A**	0.62	**I**	**A**	**I**	**AI(S:AI/V:AII/DR:AI/R:AI/AC:BII)**
3,5-Dihydroxybenzoic acid	Rye	Pass	A	II	0.60	A	II	−0.33	A	I	0.80	A	I	1.00	B	II	0.60	**A**	0.52	**II**	**A**	**I-II**	**AII(S:AII/V:AII/DR:AI/R:AI/AC:BII)**
3,5-Dihydroxybenzoic acid	Wheat	Pass	A	II	0.60	A	I	−0.17	C	IV	0.20	B	I	0.80	B	II	0.60	**B**	0.38	**II**	**A**	**I-II**	**BII(S:AII/V:AI/DR:CIV/R:BI/AC:BII)**
Atractyligenin glucuronide	Coffee	Pass	C	I	1.00	C	II	−0.50	C	IV	0.20	B	I	0.80	B	IV	0.00	**C**	0.35	**II**	**B-C**	**II**	**CII(S:CI/V:CII/DR:CIV/R:BI/AC:BIV)**
Cyclo(L-leu-L-pro)	Coffee	Pass	C	II	0.40	D	I	0.00	C	IV	0.20	C	II	0.40	B	IV	0.00	**C**	0.23	**III**	**C**	**II**	**CIII(S:CII/V:DI/DR:CIV/R:CII/AC:BIV)**
Caffeine	Coffee	Pass	A	IV	0.00	C	II	−0.33	B	II	0.60	B	I	1.00	B	II	0.60	**B**	0.34	**II**	**C**	**II**	**BII(S:AIV/V:CII/DR:BII/R:BI/AC:BII)**
1-Methylxanthine	Coffee	Pass	A	II	0.60	C	I	−0.17	C	III	0.40	C	II	0.70	B	IV	0.00	**C**	0.35	**II**	**C**	**II**	**CII(S:AII/V:CI/DR:CIII/R:CII/AC:BIV)**
Trigonelline	Coffee	Pass	A	III	0.13	C	I	−0.17	C	III	0.40	B	II	0.60	B	II	0.60	**B**	0.27	**III**	**B-C**	**III**	**BIII(S:AIII/V:CI/DR:CIII/R:BII/AC:BII)**
Trigonelline	Green beans	Pass	A	IV	0.00	C	I	−0.17	D	IV	0.00	B	II	0.60	B	II	0.60	**B**	0.15	**IV**	**B-C**	**IV**	**BIV(S:AIV/V:CI/DR:DIV/R:BII/AC:BII)**
Trigonelline	Peas	Pass	A	IV	0.00	C	I	−0.17	C	II	0.60	C	IV	0.20	B	II	0.60	**C**	0.19	**IV**	**B-C**	**IV**	**CIV(S:AIV/V:CI/DR:CII/R:CIV/AC:BII)**
Hesperetin	Orange	Pass	A	II	0.60	C	II	−0.33	B	I	0.80	B	I	1.00	B	II	0.60	**B**	0.52	**II**	**A-B**	**II**	**BII(S:AII/V:CII/DR:BI/R:BI/AC:BII)**
Hesperetin 3-*O*-glucuronide	Orange	Pass	A	II	0.60	C	II	−0.33	D	IV	0.00	B	I	0.80	B	IV	0.00	**B**	0.25	**III**	**A-B**	**II**	**BIII(S:AII/V:CII/DR:DIV/R:BI/AC:BIV)**
Hesperetin 7-*O*-glucuronide	Orange	Pass	A	II	0.60	C	II	−0.33	C	IV	0.00	C	III	0.30	B	II	0.60	**C**	0.18	**IV**	**A-B**	**II**	**CIV(S:AII/V:CII/DR:CIV/R:CIII/AC:BII)**
Proline betaine	Orange	Pass	A	II	0.40	D	I	−0.17	A	I	1.00	A	I	1.00	B	II	0.60	**A**	0.56	**II**	**A**	**I**	**AII(S:AII/V:DI/DR:AI/R:AI/AC:BII)**
Proline betaine	Citrus fruits	Pass	A	II	0.60	D	I	−0.17	A	I	1.00	A	I	1.00	B	II	0.60	**A**	0.61	**I**	**A**	**I**	**AI(S:AII/V:DI/DR:AI/R:AI/AC:BII)**
Valencic acid glucuronide	Celery	Fail																					

Levels A: BFIs with profound validation data available covering most aspects for full validation, Level B: BFIs with good data availability, Level C: BFIs with fair data availability, Level D: BFIs that only have the minimum required data available, e.g., newly discovered BFI candidates. Level I: BFIs with excellent quality, Level II: BFIs with good quality, Level III: BFIs with fair quality, Level IV: BFIs with poor quality; *measured as aglycon after conjugate hydrolysis; **including apple-derived foods.

Application and evaluation of the FoodPhyt scoring system

To demonstrate the ability of the developed scoring system to adequately evaluate any BFI, it was applied to a selection of 27 BFI-food pairs of different levels of validity. Table 2 summarizes the results for the pass-or-fail criteria, the calculated sub-scores for specificity, variability, dose-response, robustness, and analytical characterization as well as the final overall score and string notation for these examples. It also presents the consensus opinion by the FoodPhyt consortium (expert view) on data availability and BFI quality that allowed the evaluation of the newly developed and calculated score. The detailed underlying validation data, with the original references, are provided in Supplemental Material II in Tables S2–S4.

Overall, for most of the 27 BFI-food pairs, our data availability scores agree with consortium consensus. The few mismatches (e.g., *N*-caproylhistamine for tomato, phloretin and phloretin 2-*O*-glucuronide for apple, hesperetin 7-*O*-glucuronide for orange) mainly concern variability and/or dose-response, underscoring the need to evaluate these criteria carefully. For the BFI quality, scores match expert opinion for half of the 27 BFI-food pairs. If they differed, the scores were lower, reflecting limited data availability and missing studies on key validation criteria (variability in the food, intra-/inter-individual variability, dose-response with ≥3 doses, and validated quantitative analytical methods). Only three BFI show quality scores two or more levels below the FoodPhyt consensus, explained by low data availability (hesperetin 7-*O*-glucuronide) or high variability in the food (SFN, SFN-NAC). The sub-score strings and underlying data (Tables 1 and 2) make it easy to identify discrepancies and knowledge gaps. For example, cyclo(L-Leu-L-Pro) for coffee has an overall rating of CIII (0.25), while the consortium's suggestion was CII. Its sub-score string CIII(S:CII/V:DI/DR:CIV/R:CII/AC:BIV) as well as the underlying data reveal (i) low data-availability, (ii) dose-response poorly demonstrated in two observational studies with weak to moderate evidence, (iii) robustness supported by three observational but no controlled intervention studies, and (iv) the absence of validated quantification methods. This example illustrates how the scoring system can quickly identify specific validation needs for each BFI.

Four of the selected BFIs, ergothioneine, DHBA, SFN-NAC, and proline betaine, were used to test whether the scoring scheme can differentiate a BFI for a single *versus* a food group. In this scenario, the BFI content typically varies in other foods of the same food group. Targeting a single food or a food group has a large impact on the specificity sub-score. Effectively, SFN-NAC has a specificity sub-score of 0.40 when associated with broccoli, compared to 1.00 for *Brassica* vegetables. The same applies to mushrooms *versus* champignons for ergothioneine, whole grains *versus* rye, wheat and barley for DHBA, and citrus fruits *versus* orange, lemon, lime and grapefruit for proline betaine.

By testing different scenarios, (i) scoring BFIs of different quality, (ii) scoring BFIs with different data availability, and (iii) scoring BFIs with different associations to a single food *versus* a food group, against expert opinions by the FoodPhyt consortium, the scoring system demonstrated a robust and consistent framework for assessing BFIs validity.

For each BFI, the evaluation of the validation data (as described in the previous sections for each validation criterion), should be systematically documented and made available along with the calculated score. This practice allows other researchers to verify the decisions made during the evaluation process, and facilitate the score update as soon as new data become available. A template for extraction of validation data, similar to Table 1, is provided in the Supplemental Material II (Supplemental Table S5) to serve as a guidance for an easy application of the scoring system and thus to ensure a harmonized and transparent scoring of BFIs. With notes and predefined drop-down menus, it is designed to facilitate the tracking of all relevant information required to perform a BFI scoring.

The reliability of the scoring depends on the quality of the underlying validation studies. An example is related to the evaluation of phloridzin specificity for apples (see Supplemental Material II Table S3). One study reported a higher content of phloridzin in pomegranate juice compared to apple juice, resulting in a factor $S_{content} = 0$. However, during discussions with our multidisciplinary working group with experts in (poly)phenol research, it appeared that the compound described in pomegranate juice was misidentified as phloridzin based on an inappropriate method used in that study. Therefore, it is advisable to double-check data that have a strong influence on the scoring of a BFI.

A total of 16 global validation scores can be achieved, as shown in a matrix in Figure 8. Using the colors and symbols of this matrix, users can quickly gauge the recommended applicability of the associated BFI. BFIs scored as "ready for routine use" have a solid data basis and only minor, well-characterized limitations, making them suitable for ranking intake and, when calibration equations are available and inter-individual variation is low, potentially also for predicting intake. BFIs "suitable for exploratory or compliance work or possibly in BFI panels" have clear validation gaps. While their applicability may improve as additional data becomes available, they are currently best suited for distinguishing consumers from non-consumers (compliance), for exploratory intake ranking or as part of a multi-marker BFI panel. BFIs classified as "requires additional validation data" are currently supported by limited evidence, and their applicability remains uncertain until further research updates their scores. BFIs classified as "not recommended" have failed too many validation criteria to be considered usable in practice and are unlikely to become suitable in the future. Comparing these 4 groups with the utility scale by Cuparencu et al. (2024) reveals strong consistency: the "ready for routine use"-group corresponds to utility level 1; the "suitable for exploratory or compliance work or possibly in BFI panels" matches the "c" designation for combinatory use of BFIs; and the "requires additional validation data" group aligns with utility level 2 or 3 depending on which validation criteria are missing or those BFIs were not investigated by Cuparencu et al. (2024).

The overall FoodPhyt score gives a first indication of applicability, but the suggested categories are broad and may overlap. Low overall scores can result either from uniformly low sub-scores or a single very low sub-score. Therefore, users are encouraged to inspect the detailed string of sub-scores, to identify potential limitations or issues for the

Figure 8. Matrix showing the different levels of the overall bipartite BFI validation score. The horizontal axis enumerates the BFI Quality score (I: excellent; II: good; III: moderate; IV: poor), while the vertical axis lists the BFI Data availability score (A: comprehensive; B: good; C: limited; D: minimal) in descending order. Four color classes and symbols encode the overall recommendation for applicability (see legend). Some cells contain half-filled squares (triangles) to indicate overlapping recommendations.

intended application, as well as missing information with respect to particular validation criteria.

The BFI-Hub – a searchable online inventory for BFIs

As BFI validation data is scattered across diverse research areas and varies in quality, there is a need for a centralized structured platform to organize and interpret the growing literature data on BFIs. Such a resource can not only improve access to BFI validation data but can also ensure that evaluation results are clearly presented according to established guidelines. To address these needs, we developed BFI-Hub, a new searchable database specific for BFIs, through a collaboration between the JPI FoodPhyt (https://foodphyt.hub.inrae.fr/) and PlantIntake projects (https://www.healthydietforhealthylife.eu/project/plantintake). The BFI-Hub database is accessible here: https://biomarker.plantintake.eu/app/. Screenshots of the webpage are provided in Supplemental Material I (see Supplemental Figures S11–S16).

BFI-Hub provides information on BFI scores and associated validation data under the tab "Biomarker validity" (see Supplemental Figure S12). The "Scored Biomarkers"-tab, displays both the sub-scores, the overall score and the respective string, offering users a comprehensive and quick overview. With the provided filter and sorting options, users can easily access the information relevant to their specific use case. The validation data comprise plant and food related data along with human study related information, displayed in five separate tables (see Supplemental Figure S12) corresponding to each validation criterion: (i) The "Metabolism"-tab contains the known BFI precursors, information on the metabolic origin of the BFI (e.g., host metabolism, or gut microbiota), and its known co-metabolites. (ii) The "Specificity"-tab lists known sources of the BFI, provides a short study description, and a comparison of BFI contents to the food of interest. (iii) The "Variability"-tab contains descriptions about potential factors that may cause changes in the BFI content in specific foods. Furthermore, it lists information regarding variability factors in biological samples. (iv) The "Dose-response & Robustness"-tab lists data from human studies where the BFI was discovered/investigated. These data include basic information about the study (such as study type, number of participants, cohort characteristics), a description of the associated food, along with time- and dose-response related data. (v) The "Analytical characterization"-tab provides characteristics of the applied method to measure the BFI in biological samples, including instrumentation, sample preparation, validation status, and MSI level.

BFI-Hub will serve as a repository for BFIs, integrating our scoring system to evaluate the validity of existing and newly discovered biomarkers. By disseminating the BFI scoring system through this database, we aim to promote its broad application across the scientific community. The database is intended to evolve with the progress in the field, and the possibility is offered to other researchers to contribute to the curation of BFI-Hub. New BFI candidates that meet the pass-or-fail criteria can be proposed through submission of a completed template, which will be reviewed by the database curators. Additional contributions to data curation regarding the level of BFI validation are highly desired. To support this, guidelines and templates are provided here in Supplemental Table S6, and additional templates are available through the website, to assist researchers in preparing their data for submission. We expect that BFI-Hub will not

only enhance the comparability of BFIs but also foster their use for dietary assessment in future studies.

Discussion

A new BFI scoring system was developed, building on prior community efforts, through an iterative, consensus-driven process by researchers with complementary expertise. Merging multidisciplinary perspectives should enhance the adaptability and overall acceptance of the FoodPhyt scoring system to advance food and health research.

Scoring system concept

A key originality of our scoring system is that it was designed to assess both BFI quality and data availability for each validation criterion, expressed through a set of sub-scores. The data availability facilitates the identification of missing validation data and reflects the trustworthiness of the BFI quality score.

Additionally, the new scoring system can detect subtle differences in BFI quality through a point-based approach. This method calculates individual sub-scores for each validation criterion, which are subsequently combined to determine the overall score. The examples given for each validation criterion (see respective sections) demonstrate the improved differentiation in BFI quality achieved with a point-based scoring system compared to a categorial system, such as with the four-level utility scale recently proposed by Cuparencu et al. (2024). Both types of systems use validation criteria as the foundation for scoring but differ in application. Beyond incorporating data availability and strong BFI differentiation capabilities, our scoring system includes an additional mechanism that ensures a high level of BFI quality and trustworthiness – namely, the pass-or-fail criteria. BFIs having an ambiguous level of identification or an ambiguous relation with the food they are supposed to reflect (plausibility criteria), should not undergo unnecessary validation efforts or being disseminated until a minimum set of criteria have been met.

This scoring system was developed on a comprehensive data foundation. Literature data from human studies were gathered through a systematic literature search conducted in accordance with the Guidelines for Biomarker of Food Intake Reviews (Praticò et al. 2018). Additional information specific to each food item required targeted literature searches, as detailed in the Supplemental Material I. A thorough search for high-quality literature data for each validation criterion was mandatory to collect comprehensive datasets. The use of standardized data collection protocols and recommendations ensures consistency and comparability across studies.

In developing this scoring system, we tried to balance ease of application with the need of expertise in the field to evaluate the correctness of underlying data, and therefore, the score itself. The scoring results should not differ by more than one level between two researchers, otherwise reasons for discrepancy must be explained and the BFI should be used with caution. Overall, our scoring system provided more conservative estimates on validity of BFIs than reported consensus opinions of the FoodPhyt consortium.

Implementing a bipartite score (BFI quality and data availability) with four levels enables a differentiated assessment of the BFI's strengths and weaknesses. Knowing the performance of a BFI and the utility level is crucial to select a suitable BFI for different types of application, which are well discussed by Cuparencu et al. (2024). For compliance, a BFI with limited robustness is sufficient, whereas for food intake assessment, highly specific BFIs are necessary. Besides the validity of a BFI, additional information related to its application might be collected, such as potential population-related constraints, availability of standards, or pharmacokinetic data. Pharmacokinetic data relates to practical applicability by indicating the appropriate sampling time and whether the marker reflects short-term (recent) or long-term (habitual) intake. Information about the kinetic parameters of a BFI, such as peak concentration (c_{max}), peak time (t_{max}), half-life time ($t_{1/2}$), and excretion time should be collected in the template for all BFIs (see Supplementary Material II; Supplementary Table S4). However, numeric thresholds and validity ranking methods are difficult to established for kinetics parameters, and would differ for short- and long-term markers. Consequently, pharmacokinetic data were retained as complementary informative data that are important for applying BFIs in practice, but were not scored in the FoodPhyt scoring system.

Additionally, the scoring system is integrated into a user-friendly database (BFI-Hub), allowing researchers to easily obtain BFI validity scores. As the field of BFIs is continuously developing, an ongoing community effort is necessary to add new BFIs and update the validation data and scoring itself. The dissemination of the scoring system through the database facilitates the adoption of the scoring system to set a new standard for BFI validation.

Limitations

The selection of scoring criterion weights and the definition of threshold values relied primarily on consensus judgment and empirical observation rather than on a fully formal, data-driven optimization. Nevertheless, the FoodPhyt consortium assembled a large body of validation data, examined the underlying distribution of scores and how these fit with explainable thresholds (e.g., negative values generated by penalization as poorest quality), and performed systematic sensitivity analyses to assess how alternative thresholds and weightings would affect the final score and how resulting scores fit with other frameworks such as the 4-level utility scale (Cuparencu et al. 2024). Despite these efforts many parameters still reflect a degree of subjectivity that may necessitate adjustments of weights and thresholds once a larger body of BFIs has been scored.

So far, the framework has been developed and tested only on BFIs for plant-based foods. Although we expect the methodology to be transferable to animal-derived matrices, minor issues may have been missed and could require modest adjustments in future.

Thorough validation of BFIs requires extensive literature search and data analysis, making the scoring system time-consuming and potentially limiting its adoption. The

integration of AI-assisted tools for literature screening and data extraction holds promise to significantly reduce workload of extracting the growing number of emerging BFI studies and integrating these studies in new assessments of BFI validity, particularly when validating a large number of BFIs.

The applicability of some criteria might be imperfect due to limited data availability for the majority of BFIs. While the selection of BFI or validation data for the scoring system was not limited to Western populations, there may be a bias due to the scarcity of validation data in other populations. This hinders the evaluation of population-dependent responses of a BFI, an aspect of inter-individual variability. Nonetheless, the scoring system can universally score BFIs in any population, and future validation data across diverse groups might be added for evaluation of both population-independent or population-dependent BFIs. Due to the limited availability of data, important aspects – such as the detectable range of food intake (i.e., the minimum amount required for measurement in biological samples and the maximum amount at which saturation effects may occur) – could not be included in our scoring system but may be incorporated in the future. Such additional aspects can be important depending on the intended use of the BFI (e.g., as a quantitative BFI). If new data that affect the scoring system emerges, the community should discuss necessary improvements to create an updated version.

For many foods, the availability of highly specific biomarkers is limited. A combination of two or more semi-specific BFIs is a way to improve the specificity in comparison to a single BFI. Such combined BFI panels are a promising step forward, but require careful validation (Garcia-Aloy et al. 2014; Rothwell et al. 2014; Scalbert et al. 2014; Zhu et al. 2016; Vazquez-Manjarrez et al. 2019). The current version of the scoring system is not suitable to assess the validation of combined panels of BFIs. Key validation criteria such as specificity, dose-response, and robustness require dedicated studies investigating the panel as a defined set of BFIs. Even though the investigation and application of combined BFI panels undergoes continuous evolution at the moment (Chakraborty et al. 2025), such studies are still scarce and essential data are missing to enable scoring of BFI panels. Therefore, this should be considered as a perspective for future refinement and adjustment of the current system, once more comprehensive validation data for BFI panels becomes available.

Conclusion

The validation of BFIs is a crucial step toward their routine application in nutrition and health research as well as broader use in public health practice. The FoodPhyt scoring system is the first validation system to systematically assess both BFI quality and data availability, providing a direct indication of a score's trustworthiness while also highlighting potential data gaps. It is a quantitative, guided scoring system with detailed recommendations, developed by a multidisciplinary consortium, to promote rigorous and harmonized application of the BFI scoring system. The FoodPhyt validation system will enable the comparison and selection of an increasing number of BFIs with various levels of validity, in full awareness of the current state of evidence for specific applications. The new searchable BFI-Hub database (https://biomarker.plantintake.eu/app/), integrating the BFI scoring system, will enable scientists to easily select the most suitable BFI for nutrition and health research.

The presented scoring system is intended to evolve through community engagement and constructive feedback. Wider application of the scoring system to a large number of BFIs will help to refine the assumptions underlying the scoring algorithm and guide improvements in future versions.

Future efforts should focus on the validation of BFIs across diverse populations with systematic statistical analysis to support their relevance and adoption beyond Western settings. Advances in high-throughput analytical platforms will allow rapid, accurate measurement of large biomarker panels and their integration with questionnaires and digital tools to improve dietary assessment. In this setting, a harmonized, community-endorsed scoring system for BFI validity will be essential to select appropriate biomarkers and ensure consistent and transparent evaluation. Taken together, validated BFIs and self-reported dietary intake will markedly improve dietary assessment and reduce uncertainty in diet-health associations.

Acknowledgements

We thank David Fuentes from the Max Rubner-Institute for scoring the BFIs in the scoring system evaluation process as an independent person, as he was not involved in the development process.

Authors' contributions

The authors' contribution to this work were as follows: CH, CIM, SEK, CM: designed the research; CH, CIM, SEK: are responsible for the conception and the realization of the scoring system; CH, CIM, CF, RE, RLR, RL, PM, CoM, SN, ARM, MS, FTB, DSW, CM, and SEK contributed to the development of the scoring system in the consensus-based process; CH: screened the literature and conducted the data extraction and analysis; CH, CF: performed the BFI scoring; GK, LV, SB, CH: developed the online database BFI-Hub; CH, CIM: drafted the paper; All authors critically reviewed the manuscript and approved the final version. All authors agree to be accountable for all aspects of the work.

Disclosure statement

The authors declare no conflicts of interest.

Funding

This work has received funding from the French Agency of National Research (ANR, grant 19-HDH2-0002-01), the Ministero delle Politiche Agricole Alimentari e Forestali (MIPAAF, ID 1160), the Spanish Ministry of Science, Innovation, and Universities (AC19/00100), German Federal Ministry of Agriculture, Food and Regional Identity (BMLEH). through the Federal Office for Agriculture and Food (BLE) (2819ERA11F), the Federal Ministry of Education and Research (BMBF), represented by the Project Management Agency in the German Aerospace Center (DLR-PT) (01EA2204A and 01EA2204B), and the Swedish Research Council for Sustainable Development as part of the FoodPhyt project, under the umbrella of the European Joint Programming Initiative "A Healthy Diet for a Healthy Life" (JPI HDHL) (2019–02201) and of the ERA-NET Cofund HDHL INTIMIC (GA N° 727565 of the EU Horizon 2020 Research and Innovation Programme). GK received funding from the National Institutes of Health (NIH)/ National Institute on Aging (NIA) through the grant U19AG063744 (Alzheimer's Gut Microbiome Project).

ORCID

Christoph Hassenberg http://orcid.org/0000-0002-5946-8623
Carina I. Mack http://orcid.org/0000-0001-8617-7947
Claudia Favari http://orcid.org/0000-0003-0669-2223
Sylvie Baier http://orcid.org/0000-0002-9120-6279
Ramon Estruch http://orcid.org/0000-0003-1260-4445
Gabi Kastenmüller http://orcid.org/0000-0002-2368-7322
Rosa Lamuela-Raventós http://orcid.org/0000-0002-1287-4560
Rikard Landberg http://orcid.org/0000-0002-6399-7608
Pedro Mena http://orcid.org/0000-0003-2150-2977
Costanza Michelini http://orcid.org/0000-0002-5389-6141
Stefania Noerman http://orcid.org/0000-0003-4615-7268
Ana Rodriguez-Mateos http://orcid.org/0000-0003-3242-402X
Mia Stråvik http://orcid.org/0000-0001-9651-6754
Francisco A. Tomás-Barberán http://orcid.org/0000-0002-0790-1739
Lara Vehovec http://orcid.org/0009-0000-0613-6820
David S. Wishart http://orcid.org/0000-0002-3207-2434
Claudine Manach http://orcid.org/0000-0003-1094-3363
Sabine E. Kulling http://orcid.org/0000-0002-9956-2702

Abbreviations

ADME	absorption, distribution, metabolism and excretion
BFIs	biomarkers of food intake
CAWG	chemical analysis working group
DHBA	3,5-dihydroxybenzoic acid
DR	dietary records
EFSA	European Food Safety Authority
EMA	European Medicines Agency
FDA	Food and Drug Administration
FFQ	food frequency questionnaires
GSTM1	glutathione-S-transferase M1
HDHL	Healthy Diet for a Healthy Life
HMDB	human metabolome database
HRMS	high-resolution mass spectrometry
ICC	intra-class correlation coefficient
ICH	International Council for Harmonisation
InChI	International Chemical Identifier
IUPAC	International Union of Pure and Applied Chemistry
JPI	Joint Programming Initiative
LOD	limit of detection
LOQ	limit of quantification
MS	mass spectrometer
MSI	Metabolomics Standards Initiative
NMR	nuclear magnetic resonance spectroscopy
SFN	sulforaphane
SFN-NAC	sulforaphane-*N*-acetyl cysteine

References

Abreu TC et al. 2021. Validity coefficient of repeated measurements of urinary marker of sugar intake is comparable to urinary nitrogen as marker of protein intake in free-living subjects. Cancer Epidemiol Biomarkers Prev. 30(1):193–202. https://www.ncbi.nlm.nih.gov/pubmed/32998945. https://doi.org/10.1158/1055-9965.EPI-20-0271

Al Janobi AA et al. 2006. Quantitative measurement of sulforaphane, iberin and their mercapturic acid pathway metabolites in human plasma and urine using liquid chromatography-tandem electrospray ionisation mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci. 844(2):223–234. https://www.ncbi.nlm.nih.gov/pubmed/16931178. https://doi.org/10.1016/j.jchromb.2006.07.007

Andersen MB et al. 2013. Discovery of exposure markers in urine for Brassica-containing meals served with different protein sources by UPLC-qTOF-MS untargeted metabolomics. Metabolomics. 9(5):984–997. https://doi.org/10.1007/s11306-013-0522-0

Andersen MB et al. 2014. Discovery and validation of urinary exposure markers for different plant foods by untargeted metabolomics. Anal Bioanal Chem. 406(7):1829–1844. https://www.ncbi.nlm.nih.gov/pubmed/24390407; https://link.springer.com/content/pdf/10.1007/s00216-013-7498-5.pdf. https://doi.org/10.1007/s00216-013-7498-5

Angelino D, Jeffery E. 2014. Glucosinolate hydrolysis and bioavailability of resulting isothiocyanates: focus on glucoraphanin. J Funct Foods. 7:67–76. https://doi.org/10.1016/j.jff.2013.09.029

Atkinson W, Downer P, Lever M, Chambers ST, George PM. 2007. Effects of orange juice and proline betaine on glycine betaine and homocysteine in healthy male subjects. Eur J Nutr. 46(8):446–452. https://link.springer.com/content/pdf/10.1007/s00394-007-0684-5.pdf. https://doi.org/10.1007/s00394-007-0684-5

Baenas N, Suárez-Martínez C, García-Viguera C, Moreno DA. 2017. Bioavailability and new biomarkers of cruciferous sprouts consumption. Food Res Int. 100(Pt 1):497–503. https://doi.org/10.1016/j.foodres.2017.07.049

Beckmann M et al. 2020. A standardized strategy for simultaneous quantification of urine metabolites to validate development of a biomarker panel allowing comprehensive assessment of dietary exposure. Mol Nutr Food Res. 64(20):e2000517. https://doi.org/10.1002/mnfr.202000517

Biondi F et al. 2021. Environmental conditions and agronomical factors influencing the levels of phytochemicals in brassica vegetables responsible for nutritional and sensorial properties. Appl Sci. 11(4):1927. https://doi.org/10.3390/app11041927

Borthwick AD, Da Costa NC. 2017. 2,5-diketopiperazines in food and beverages: taste and bioactivity. Crit Rev Food Sci Nutr. 57(4):718–742. https://www.ncbi.nlm.nih.gov/pubmed/25629623. https://doi.org/10.1080/10408398.2014.911142

Bouranis JA et al. 2021. Composition of the gut microbiome influences production of sulforaphane-nitrile and iberin-nitrile from glucosinolates in broccoli sprouts. Nutrients. 13(9):3013. https://doi.org/10.3390/nu13093013

Brennan L, Hu FB. 2019. Metabolomics-based dietary biomarkers in nutritional epidemiology-current status and future opportunities. Mol Nutr Food Res. 63(1):e1701064. https://www.ncbi.nlm.nih.gov/pubmed/29688616. https://doi.org/10.1002/mnfr.201701064

Brennan L. 2017. Metabolomics: a tool to aid dietary assessment in nutrition. Curr Opin Food Sci. 16:96–99. https://doi.org/10.1016/j.cofs.2017.09.003

Brouwer-Brolsma EM, Brandl B, Buso MEC, Skurk T, Manach C. 2020. Food intake biomarkers for green leafy vegetables, bulb vegetables, and stem vegetables: a review. Genes Nutr. 15(1):7. https://genesandnutrition.biomedcentral.com/track/pdf/10.1186/s12263-020-00667-z. https://doi.org/10.1186/s12263-020-00667-z

Cabezas J, Lucey MR, Bataller R. 2016. Biomarkers for monitoring alcohol use. Clin Liver Dis (Hoboken). 8(3):59–63. https://doi.org/10.1002/CLD.571

Chakraborty H et al. 2025. The dietary biomarkers development consortium: an initiative for discovery and validation of dietary biomarkers for precision nutrition. Curr Dev Nutr. 9(5):107435. https://doi.org/10.1016/j.cdnut.2025.107435

Charron CS et al. 2018. Absorption and metabolism of isothiocyanates formed from broccoli glucosinolates: effects of BMI and daily consumption in a randomised clinical trial. Br J Nutr. 120(12):1370–1379. https://doi.org/10.1017/s0007114518002921

Clarke ED, Rollo ME, Pezdirc K, Collins CE, Haslam RL. 2020. Urinary biomarkers of dietary intake: a review. Nutr Rev. 78(5):364–381. https://doi.org/10.1093/nutrit/nuz048

Clarke JD et al. 2011. Comparison of isothiocyanate metabolite levels and histone deacetylase activity in human subjects consuming broccoli sprouts or broccoli supplement. J Agric Food Chem. 59(20):10955–10963. https://pubs.acs.org/doi/pdf/10.1021/jf202887c. https://doi.org/10.1021/jf202887c

Committee for Medicinal Products for Human Use, European Medicines Agency. 2011. "Guideline on bioanalytical method validation." https://www.ema.europa.eu/en/bioanalytical-method-validation-scientific-guideline

Conaway CC et al. 2000. Disposition of glucosinolates and sulforaphane in humans after ingestion of steamed and fresh broccoli. Nutr Cancer. 38(2):168–178. https://doi.org/10.1207/s15327914nc382_5

Costa-Perez A et al. 2023. Systematic review on the metabolic interest of glucosinolates and their bioactive derivatives for human health. Nutrients. 15(6):1424. https://www.ncbi.nlm.nih.gov/pubmed/36986155; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10058295/pdf/nutrients-15-01424.pdf. https://doi.org/10.3390/nu15061424

Creek DJ et al. 2014. Metabolite identification: are you sure? And how do your peers gauge your confidence? Metabolomics. 10(3):350–353. https://doi.org/10.1007/s11306-014-0656-8

Cui Q et al. 2021. A meta-analysis of the reproducibility of food frequency questionnaires in nutritional epidemiological studies. Int J Behav Nutr Phys Act. 18(1):12. . https://doi.org/10.1186/s12966-020-01078-4

Cuparencu C et al. 2019. Biomarkers of meat and seafood intake: an extensive literature review. Genes Nutr. 14(1):35. https://www.ncbi.nlm.nih.gov/pubmed/31908682. https://doi.org/10.1186/s12263-019-0656-4

Cuparencu C et al. 2024. Towards nutrition with precision: unlocking biomarkers as dietary assessment tools. Nat Metab. 6(8):1438–1453. https://doi.org/10.1038/s42255-024-01067-y

D'Angelo S et al. 2019. Combining biomarker and food intake data: calibration equations for citrus intake. Am J Clin Nutr. 110(4):977–983. https://doi.org/10.1093/ajcn/nqz168https://www.ncbi.nlm.nih.gov/pubmed/31432078

de la O V et al. 2025. Exhaustive search of dietary intake biomarkers as objective tools for personalized nutrimetabolomics and precision nutrition implementation. Nutr Rev. 83(5):925–942. https://doi.org/10.1093/nutrit/nuae133

Deng Y et al. 1997. Assay for the (R)- and (S)-enantiomers of salsolinols in biological samples and foods with ion-pair high-performance liquid chromatography using beta-cyclodextrin as a chiral mobile phase additive. J Chromatogr B Biomed Sci Appl. 689(2):313–320. https://doi.org/10.1016/S0378-4347(96)00359-3https://www.ncbi.nlm.nih.gov/pubmed/9080316.

Dominguez-Perles R et al. 2014. A new ultra-rapid UHPLC/MS/MS method for assessing glucoraphanin and sulforaphane bioavailability in human urine. Food Chem. 143:132–138. https://doi.org/10.1016/j.foodchem.2013.07.116

Dragsted LO et al. 2018. Validation of biomarkers of food intake-critical assessment of candidate biomarkers. Genes Nutr. 13(1):14. https://doi.org/10.1186/s12263-018-0603-9

Duncan MW, Smythe GA, Nicholson MV, Clezy PS. 1984. Comparison of high-performance liquid chromatography with electrochemical detection and gas chromatography-mass fragmentography for the assay of salsolinol, dopamine and dopamine metabolites in food and beverage samples. J Chromatogr. 336(1):199–209. https://doi.org/10.1016/s0378-4347(00)85142-7.

Eisele TA, Drake SR. 2005. The partial compositional characteristics of apple juice from 175 apple varieties. J Food Compos Anal. 18(2–3):213–221. https://doi.org/10.1016/j.jfca.2004.01.002

French CD et al. 2023. Assessing repeated urinary proline betaine measures as a biomarker of usual citrus intake during pregnancy: sources of within-person variation and correlation with reported intake. Metabolites. 13(8):904. https://doi.org/10.3390/metabo13080904

Gallardo-Gómez D, Richardson R, Dwan K. 2024. Standardized mean differences in meta-analysis: a tutorial. Cochrane Evid Synth Methods. 2(3):e12047. https://doi.org/10.1002/cesm.12047

García CJ, Beltrán D, Tomás-Barberán FA. 2020. Human gut microbiota metabolism of dietary sesquiterpene lactones: untargeted metabolomics study of lactucopicrin and lactucin conversion in vitro and in vivo. Mol Nutr Food Res. 64(21):e2000619. https://doi.org/10.1002/mnfr.202000619

Garcia-Aloy M et al. 2014. Novel multimetabolite prediction of walnut consumption by a urinary biomarker model in a free-living population: the PREDIMED study. J Proteome Res. 13(7):3476–3483. https://www.ncbi.nlm.nih.gov/pubmed/24882253. https://doi.org/10.1021/pr500425r

Garcia-Aloy M et al. 2017. Novel strategies for improving dietary exposure assessment: multiple-data fusion is a more accurate measure than the traditional single-biomarker approach. Trends Food Sci Technol. 69:220–229. https://doi.org/10.1016/j.tifs.2017.04.013

Garcia-Aloy M et al. 2019. Biomarkers of food intake for nuts and vegetable oils: an extensive literature search. Genes Nutr. 14(1):7. https://doi.org/10.1186/s12263-019-0628-8

Garcia-Aloy M et al. 2020. Discovery of intake biomarkers of lentils, chickpeas, and white beans by untargeted lc-ms metabolomics in serum and urine. Mol Nutr Food Res. 64(13):e1901137. https://doi.org/10.1002/mnfr.201901137

Gasper AV et al. 2005. Glutathione S-transferase M1 polymorphism and metabolism of sulforaphane from standard and high-glucosinolate broccoli. Am J Clin Nutr. 82(6):1283–1291. https://www.ncbi.nlm.nih.gov/pubmed/16332662. https://doi.org/10.1093/ajcn/82.6.1283

Genghof DS. 1970. Biosynthesis of ergothioneine and hercynine by fungi and actinomycetales. J Bacteriol. 103(2):475–478. . https://doi.org/10.1128/jb.103.2.475-478.1970

Gibbons H et al. 2017. Demonstration of the utility of biomarkers for dietary intake assessment; proline betaine as an example. Mol Nutr Food Res. 61(10):1700037. https://doi.org/10.1002/mnfr.201700037

Gormley IC, D'Angelo S, Brennan L. 2022. Combining biomarker and food intake data. Wiley.

Guertin KA et al. 2015. Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. Am J Clin Nutr. 101(5):1000–1011. https://www.ncbi.nlm.nih.gov/pubmed/25762808. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4409687/pdf/ajcn096099.pdf. https://doi.org/10.3945/ajcn.114.096099

Hassenberg C et al. 2026. Method development and validation for quantitative determination of urinary biomarkers of food intake for multiple foods. J Chromatogr B Analyt Technol Biomed Life Sci. 1268:124793. https://www.sciencedirect.com/science/article/pii/S1570023225003472. https://doi.org/10.1016/j.jchromb.2025.124793

Hedrick VE et al. 2012. Dietary biomarkers: advances, limitations and future directions. Nutr J. 11(1):109. . https://doi.org/10.1186/1475-2891-11-109

Heinzmann SS et al. 2010. Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. Am J Clin Nutr. 92(2):436–443. https://www.ncbi.nlm.nih.gov/pubmed/20573794. https://doi.org/10.3945/ajcn.2010.29672

Heller S, McNaught A, Stein S, Tchekhovskoi D, Pletnev I. 2013. InChI – the worldwide chemical structure identifier standard. J Cheminform. 5(1):7. https://www.ncbi.nlm.nih.gov/pubmed/23343401. https://doi.org/10.1186/1758-2946-5-7

Hipolito L, Sánchez-Catalán MJ, Martí-Prats L, Granero L, Polache A. 2012. Revisiting the controversial role of salsolinol in the neurobiological effects of ethanol: old and new vistas. Neurosci Biobehav Rev. 36(1):362–378. https://www.ncbi.nlm.nih.gov/pubmed/21802444. https://doi.org/10.1016/j.neubiorev.2011.07.007

Hövelmann Y, Lewin L, Steinert K, Hübner F, Humpf HU. 2020. Mass spectrometry-based analysis of urinary biomarkers for dietary tomato intake. Mol Nutr Food Res. 64(12):e2000011. https://doi.org/10.1002/mnfr.202000011

Hu Y et al. 2024. Calibration of citrus intake assessed by food frequency questionnaires using urinary proline betaine in an observational study setting. Am J Clin Nutr. 120(1):178–186. https://doi.org/10.1016/j.ajcnut.2024.05.011

Huser S et al. 2018. Effects of isoflavones on breast tissue and the thyroid hormone system in humans: a comprehensive safety evaluation. Arch Toxicol. 92(9):2703–2748. https://doi.org/10.1007/s00204-018-2279-8

International Council for Harmonisation. 2022. ICH guideline M10: bioanalytical method validation and study sample analysis. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH).

IPSOS; European Public Affairs. 2022. "Consumer Survey on Food Supplements in the EU."

Jang HH, Lee YM, Choe JS, Kwon O. 2021. Validation of soy isoflavone intake and its health effects: a review of the development of exposure biomarkers. Nutr Res Pract. 15(1):1–11. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7838478/pdf/nrp-15-1.pdf. https://doi.org/10.4162/nrp.2021.15.1.1

Jawhara M, Sørensen SB, Heitmann BL, Andersen V. 2019. Biomarkers of whole-grain and cereal-fiber intake in human studies: a systematic review of the available evidence and perspectives. Nutrients. 11(12):2994. https://doi.org/10.3390/nu11122994

Jenab M, Slimani N, Bictash M, Ferrari P, Bingham SA. 2009. Biomarkers in nutritional epidemiology: applications, needs and new horizons. Hum Genet. 125(5–6):507–525. https://www.ncbi.nlm.nih.gov/pubmed/19357868; https://link.springer.com/content/pdf/10.1007/s00439-009-0662-5.pdf. https://doi.org/10.1007/s00439-009-0662-5

Koo TK, Li MY. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 15(2):155–163. https://www.ncbi.nlm.nih.gov/pubmed/27330520. https://doi.org/10.1016/j.jcm.2016.02.012

Kruve A et al. 2015a. Tutorial review on validation of liquid chromatography-mass spectrometry methods: part I. Anal Chim Acta. 870:29–44. https://www.ncbi.nlm.nih.gov/pubmed/25819785. https://doi.org/10.1016/j.aca.2015.02.017

Kruve A et al. 2015b. Tutorial review on validation of liquid chromatography-mass spectrometry methods: part II. Anal Chim Acta. 870:8–28. https://www.ncbi.nlm.nih.gov/pubmed/25819784. https://doi.org/10.1016/j.aca.2015.02.016

Lamy E, Scholtes C, Herz C, Mersch-Sundermann V. 2011. Pharmacokinetics and pharmacodynamics of isothiocyanates. Drug Metab Rev. 43(3):387–407. https://doi.org/10.3109/03602532.2011.569551

Landberg R et al. 2019. Biomarkers of cereal food intake. Genes Nutr. 14(1):28. https://doi.org/10.1186/s12263-019-0651-9

Landberg R et al. 2023. Dietary biomarkers-an update on their validity and applicability in epidemiological studies. Nutr Rev. 82(9):1260–1280. https://doi.org/10.1093/nutrit/nuad119https://www.ncbi.nlm.nih.gov/pubmed/37791499

Lang R et al. 2017. High-throughput quantitation of proline betaine in foods and suitability as a valid biomarker for citrus consumption. J Agric Food Chem. 65(8):1613–1619. https://doi.org/10.1021/acs.jafc.6b05824

Li KJ, Brouwer-Brolsma EM, Burton-Pimentel KJ, Vergères G, Feskens EJM. 2021. A systematic review to identify biomarkers of intake for fermented food products. Genes Nutr. 16(1):5. https://doi.org/10.1186/s12263-021-00686-4

Lloyd AJ, Beckmann M, Favé G, Mathers JC, Draper J. 2011. Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption. Br J Nutr. 106(6):812–824. https://doi.org/10.1017/s0007114511001164

Lukas B, Schmiderer C, Mitteregger U, Novak J. 2010. Arbutin in marjoram and oregano. Food Chem. 121(1):185–190. https://doi.org/10.1016/j.foodchem.2009.12.028

Lynn KS et al. 2019. Vegetable signatures derived from human urinary metabolomic data in controlled feeding studies. J Proteome Res. 18(1):159–168. https://pubs.acs.org/doi/pdf/10.1021/acs.jproteome.8b00470. https://doi.org/10.1021/acs.jproteome.8b00470

Marklund M et al. 2014. Simultaneous pharmacokinetic modeling of alkylresorcinols and their main metabolites indicates dual absorption mechanisms and enterohepatic elimination in humans. J Nutr. 144(11):1674–1680. https://doi.org/10.3945/jn.114.196220

Marks SC, Mullen W, Borges G, Crozier A. 2009. Absorption, metabolism, and excretion of cider dihydrochalcones in healthy humans and subjects with an ileostomy. J Agric Food Chem. 57(5):2009–2015. https://doi.org/10.1021/jf802757x

McNamara AE et al. 2020. Metabolomic-based approach to identify biomarkers of apple intake. Mol Nutr Food Res. 64(11):e1901158. https://doi.org/10.1002/mnfr.201901158

McNamara AE, Brennan L. 2020. Potential of food intake biomarkers in nutrition research. Proc Nutr Soc. 79(4):1–11. https://doi.org/10.1017/S0029665120007053

Melzig MF, Putscher I, Henklein P, Haber H. 2000. In vitro pharmacological activity of the tetrahydroisoquinoline salsolinol present in products from *Theobroma cacao* L. like cocoa and chocolate. J Ethnopharmacol. 73(1–2):153–159. https://doi.org/10.1016/S0378-8741(00)00291-9.

Michielsen C, Almanza-Aguilera E, Brouwer-Brolsma EM, Urpi-Sarda M, Afman LA. 2018. Biomarkers of food intake for cocoa and liquorice (products): a systematic review. Genes Nutr. 13(1):22. https://doi.org/10.1186/s12263-018-0610-x

Montonen J et al. 2012. Reliability of fasting plasma alkylresorcinol metabolites concentrations measured 4 months apart. Eur J Clin Nutr. 66(8):968–970. https://doi.org/10.1038/ejcn.2012.66

Morales D. 2022. Use of strawberry tree (*Arbutus unedo*) as a source of functional fractions with biological activities. Foods. 11(23):3838. https://doi.org/10.3390/foods11233838

Morand C et al. 2020. Why interindividual variation in response to consumption of plant food bioactives matters for future personalised nutrition. Proc Nutr Soc. 79(2):225–235. https://www.ncbi.nlm.nih.gov/pubmed/32014077. https://doi.org/10.1017/S0029665120000014

Munger LH et al. 2018. Biomarker of food intake for assessing the consumption of dairy and egg products. Genes Nutr. 13(1):26. https://www.ncbi.nlm.nih.gov/pubmed/30279743; https://genesandnutrition.biomedcentral.com/counter/pdf/10.1186/s12263-018-0615-5.pdf. https://doi.org/10.1186/s12263-018-0615-5

Musshoff F, Daldrup T, Bonte W, Leitner A, Lesch OM. 1997. Salsolinol and norsalsolinol in human urine samples. Pharmacol Biochem Behav. 58(2):545–550. https://doi.org/10.1016/S0091-3057(97)00251-7https://www.ncbi.nlm.nih.gov/pubmed/9300617.

Nahar L, Al-Groshi A, Kumar A, Sarker SD. 2022. Arbutin: occurrence in plants, and its potential as an anticancer agent. Molecules. 27(24):8786. https://doi.org/10.3390/molecules27248786

Nieman DC et al. 2015. Metabolomics-based analysis of banana and pear ingestion on exercise performance and recovery. J Proteome Res. 14(12):5367–5377. https://doi.org/10.1021/acs.jproteome.5b00909

Oliviero T, Verkerk R, Dekker M. 2018. Isothiocyanates from brassica vegetables-effects of processing, cooking, mastication, and digestion. Mol Nutr Food Res. 62(18):e1701069. https://www.ncbi.nlm.nih.gov/pubmed/29898282; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6175105/pdf/MNFR-62-na.pdf. https://doi.org/10.1002/mnfr.201701069

Ottaviani JI, Sagi-Kiss V, Schroeter H, Kuhnle GGC. 2024. Reliance on self-reports and estimated food composition data in nutrition research introduces significant bias that can only be addressed with biomarkers. Elife. 13: RP92941. . https://doi.org/10.7554/eLife.92941

Ottaviani JI, Schroeter H, Kuhnle GGC. 2023. Measuring the intake of dietary bioactives: pitfalls and how to avoid them. Mol Aspects Med. 89:101139. https://doi.org/10.1016/j.mam.2022.101139

Pallister T et al. 2016. Characterizing blood metabolomics profiles associated with self-reported food intakes in female twins. PLoS One. 11(6):e0158568. https://doi.org/10.1371/journal.pone.0158568

Parnell LD et al. 2025. Exploration of biomarkers of food intake in a Caribbean Hispanic population. Mol Nutr Food Res. 69(20):e70158. https://doi.org/10.1002/mnfr.70158

Praticò G et al. 2018. Guidelines for biomarker of food intake reviews (BFIRev): how to conduct an extensive literature search for biomarker of food intake discovery. Genes Nutr. 13(1):3. https://www.ncbi.nlm.nih.gov/pubmed/29484030; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5819202/pdf/12263_2018_Article_592.pdf. https://doi.org/10.1186/s12263-018-0592-8

Praticò G, Gao Q, Manach C, Dragsted LO. 2018. Biomarkers of food intake for Allium vegetables. Genes Nutr. 13(1):34. https://doi.org/10.1186/s12263-018-0624-4

Pujos-Guillot E et al. 2013. Mass spectrometry-based metabolomics for the discovery of biomarkers of fruit and vegetable intake: citrus fruit as a case study. J Proteome Res. 12(4):1645–1659. https://doi.org/10.1021/pr300997c

Rafiq T et al. 2021. Nutritional metabolomics and the classification of dietary biomarker candidates: a critical review. Adv Nutr. 12(6):2333–2357. https://doi.org/10.1093/advances/nmab054

Raposo F, Ibelli-Bianco C. 2020. Performance parameters for analytical method validation: controversies and discrepancies among numerous guidelines. TrAC, Trends Anal Chem. 129:115913. https://doi.org/10.1016/j.trac.2020.115913

Riedl A, Gieger C, Hauner H, Daniel H, Linseisen J. 2017. Metabotyping and its application in targeted nutrition: an overview. Br J Nutr. 117(12):1631–1644. https://doi.org/10.1017/S0007114517001611

Riggin RM, Kissinger PT. 1976. Letter: identification of salsolinol as a phenolic component in powdered cocoa and cocoa-based products. J Agric Food Chem. 24(4):900–900. https://www.ncbi.nlm.nih.gov/pubmed/956552. https://doi.org/10.1021/jf60206a043

Riggin RM, McCarthy MJ, Kissinger PT. 1976. Identification of salsolinol as a major dopamine metabolite in the banana. J Agric Food Chem. 24(1):189–191. https://www.ncbi.nlm.nih.gov/pubmed/1245664. https://doi.org/10.1021/jf60203a027

Rigutto-Farebrother J et al. 2023. Perspectives on the application of CONSORT guidelines to randomised controlled trials in nutrition. Eur J Nutr. 62(5):2319–2332. https://doi.org/10.1007/s00394-023-03137-5

Rothwell JA et al. 2014. New biomarkers of coffee consumption identified by the non-targeted metabolomic profiling of cohort study subjects. PLoS One. 9(4):e93474. https://doi.org/10.1371/journal.pone.0093474

Rothwell JA et al. 2018. Biomarkers of intake for coffee, tea, and sweetened beverages. Genes Nutr. 13(1):15. https://doi.org/10.1186/s12263-018-0607-5

Rothwell JA et al. 2019a. A metabolomic study of biomarkers of habitual coffee intake in four European countries. Mol Nutr Food Res. 63(22):e1900659. https://doi.org/10.1002/mnfr.201900659

Rothwell JA et al. 2019b. A metabolomic study of the variability of the chemical composition of commonly consumed coffee brews. Metabolites. 9(1):17. https://doi.org/10.3390/metabo9010017

Saeedi M, Khezri K, Seyed Zakaryaei A, Mohammadamini H. 2021. A comprehensive review of the therapeutic potential of alpha-arbutin. Phytother Res. 35(8):4136–4154. https://www.ncbi.nlm.nih.gov/pubmed/33724594. https://doi.org/10.1002/ptr.7076

Saenger T, Hübner F, Humpf HU. 2017. Short-term biomarkers of apple consumption. Mol Nutr Food Res. 61(3):1600629. https://doi.org/10.1002/mnfr.201600629

Saenger T, Hübner F, Lindemann V, Ganswind K, Humpf HU. 2021. Urinary biomarkers for orange juice consumption. Mol Nutr Food Res. 65(2):e2000781. https://doi.org/10.1002/mnfr.202000781

Scalbert A et al. 2014. The food metabolome: a window over dietary exposure. Am J Clin Nutr. 99(6):1286–1308. https://doi.org/10.3945/ajcn.113.076133

Schober P, Boer C, Schwarte LA. 2018. Correlation coefficients: appropriate use and interpretation. Anesth Analg. 126(5):1763–1768. https://doi.org/10.1213/ane.0000000000002864

Schymanski EL et al. 2014. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol. 48(4):2097–2098. https://www.ncbi.nlm.nih.gov/pubmed/24476540. https://doi.org/10.1021/es5002105

Shi L et al. 2020. Plasma metabolite biomarkers of boiled and filtered coffee intake and their association with type 2 diabetes risk. J Intern Med. 287(4):405–421. https://doi.org/10.1111/joim.13009

Sojo MM, Nunez-Delicado E, Sanchez-Ferrer A, Garcia-Carmona F. 2000. Oxidation of salsolinol by banana pulp polyphenol oxidase and its kinetic synergism with dopamine. J Agric Food Chem. 48(11):5543–5547. https://doi.org/10.1021/jf000293f

Sri Harsha PSC et al. 2018. Biomarkers of legume intake in human intervention and observational studies: a systematic review. Genes Nutr. 13(1):25. https://doi.org/10.1186/s12263-018-0614-6

Subar AF et al. 2015. Addressing current criticism regarding the value of self-report dietary data. J Nutr. 145(12):2639–2645. https://doi.org/10.3945/jn.115.219634

Sumner LW et al. 2007. "Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. 3(3):211–221. https://doi.org/10.1007/S11306-007-0082-2

Sun J et al. 2020. Profiling glucosinolate metabolites in human urine and plasma after broccoli consumption using non-targeted and targeted metabolomic analyses. Food Chem. 309:125660. https://doi.org/10.1016/j.foodchem.2019.125660

Tabart J, Pincemail J, Kevers C, Defraigne J, Dommes J. 2018. Processing effects on antioxidant, glucosinolate, and sulforaphane contents in broccoli and red cabbage. Eur Food Res Technol. 244(12):2085–2094. https://doi.org/10.1007/s00217-018-3126-0

Tomás-Navarro M, Navarro JL, Vallejo F, Tomás-Barberán FA. 2021. Novel urinary biomarkers of orange juice consumption, interindividual variability, and differences with processing methods. J Agric Food Chem. 69(13):4006–4017. https://doi.org/10.1021/acs.jafc.0c08144

Trost K et al. 2018. Host: microbiome co-metabolic processing of dietary polyphenols – an acute, single blinded, cross-over study with different doses of apple polyphenols in healthy subjects. Food Res Int. 112:108–128. https://doi.org/10.1016/j.foodres.2018.06.016

U.S. Department of Health and Human Services Food and Drug Administration. 2018. "Bioanalytical Method Validation – Guidance for Industry." 44. https://www.regulations.gov/document/FDA-2013-D-1020-0039

Ulaszewska CH et al. 2019. Nutrimetabolomics: an integrative action for metabolomic analyses in human nutritional studies. Mol Nutr Food Res. 63(1):e1800384. https://doi.org/10.1002/mnfr.201800384

Ulaszewska M et al. 2020. Food intake biomarkers for berries and grapes. Genes Nutr. 15(1):17. https://doi.org/10.1186/s12263-020-00675-z

Ulaszewska N et al. 2018. Food intake biomarkers for apple, pear, and stone fruit. Genes Nutr. 13(1):29. https://www.ncbi.nlm.nih.gov/pubmed/30519365; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6267079/pdf/12263_2018_Article_620.pdf. https://doi.org/10.1186/s12263-018-0620-8

Unión-Caballero A et al. 2024. Metabolome biomarkers linking dietary fibre intake with cardiometabolic effects: results from the Danish Diet, Cancer and Health-Next Generations MAX study. Food Funct. 15(3):1643–1654. https://doi.org/10.1039/D3FO04763F

Vazquez-Fresno R et al. 2019. Herbs and spices-biomarkers of intake based on human intervention studies – a systematic review. Genes Nutr. 14(1):18. https://doi.org/10.1186/s12263-019-0636-8

Vazquez-Manjarrez N et al. 2019. Discovery and validation of banana intake biomarkers using untargeted metabolomics in human intervention and cross-sectional studies. J Nutr. 149(10):1701–1713. https://doi.org/10.1093/jn/nxz125

Vazquez-Manjarrez N et al. 2020. Biomarkers of intake for tropical fruits. Genes Nutr. 15(1):11. https://doi.org/10.1186/s12263-020-00670-4

Vermeulen M, van den Berg R, Freidig AP, van Bladeren PJ, Vaes WH. 2006. Association between consumption of cruciferous vegetables and condiments and excretion in urine of isothiocyanate mercapturic acids. J Agric Food Chem. 54(15):5350–5358. https://doi.org/10.1021/jf060723n

Vermeulen M, van Rooijen HJ, Vaes WH. 2003. Analysis of isothiocyanate mercapturic acids in urine: a biomarker for cruciferous vegetable intake. J Agric Food Chem. 51(12):3554–3559. https://doi.org/10.1021/jf0341316

Wang Y et al. 2018. Untargeted metabolomics identifies novel potential biomarkers of habitual food intake in a cross-sectional study of postmenopausal women. J Nutr. 148(6):932–943. https://doi.org/10.1093/jn/nxy027

Wang Y, Hodge RA, Stevens VL, Hartman TJ, McCullough ML. 2020. Identification and reproducibility of plasma metabolomic biomarkers of habitual food intake in a us diet validation study. Metabolites. 10(10):382. https://doi.org/10.3390/metabo10100382

Weng H et al. 2020. Low oral bioavailability and partial gut microbiotic and phase II metabolism of Brussels/Witloof Chicory Sesquiterpene lactones in healthy humans. Nutrients. 12(12):3675. https://doi.org/10.3390/nu12123675

Wierzbicka R, Zamaratskaia G, Kamal-Eldin A, Landberg R. 2017. Novel urinary alkylresorcinol metabolites as biomarkers of whole grain intake in free-living Swedish adults. Mol Nutr Food Res. 61(7):1700015. https://doi.org/10.1002/mnfr.201700015

Woodside JV, Draper J, Lloyd A, McKinley MC. 2017. Use of biomarkers to assess fruit and vegetable intake. Proc Nutr Soc. 76(3):308–315. https://doi.org/10.1017/S0029665117000325

Xi M, Dragsted LO. 2019. Biomarkers of seaweed intake. Genes Nutr. 14(1):24. https://doi.org/10.1186/s12263-019-0648-4

Xiao J, Capanoglu E, Jassbi AR, Miron A. 2016. Advance on the flavonoid C-glycosides and health benefits. Crit Rev Food Sci Nutr. 56 Suppl 1: s 29–45. https://doi.org/10.1080/10408398.2015.1067595

Yuan Y et al. 2017. Metabolomic analyses of banana during postharvest senescence by (1)H-high resolution-NMR. Food Chem. 218:406–412. https://doi.org/10.1016/j.foodchem.2016.09.080

Zayed A, Sheashea M, Kassem IAA, Farag MA. 2023. Red and white cabbages: an updated comparative review of bioactives, extraction methods, processing practices, and health benefits. Crit Rev Food Sci Nutr. 63(24):7025–7042. https://doi.org/10.1080/10408398.2022.2040416

Zhou X, Gao Q, Praticò G, Chen J, Dragsted LO. 2019. Biomarkers of tuber intake. Genes Nutr. 14(1):9. https://doi.org/10.1186/s12263-019-0631-0

Zhu Y, Wang P, Sha W, Sang S. 2016. Urinary biomarkers of whole grain wheat intake identified by non-targeted and targeted metabolomics approaches. Sci Rep. 6(1):36278. https://doi.org/10.1038/srep36278