REVIEW

Improving the reporting quality of intervention trials addressing the inter-individual variability in response to the consumption of plant bioactives: quality index and recommendations

Marina Nikolic[1] · Aleksandra Konic Ristic[1,2] · Antonio González-Sarrías[3] · Geoffrey Istas[4] ·
Mireia Urpi-Sarda[5,6] · Margherita Dall'Asta[7] · Laurent-Emmanuel Monfoulet[8] · Lieselotte Cloetens[9] ·
Banu Bayram[10] · Maria Rosaria Tumolo[11] · Mihail Chervenkov[12,13] · Egeria Scoditti[14] · Marika Massaro[14] ·
Noemi Tejera[15] · Desislava Abadjieva[16] · Karen Chambers[17] · Irena Krga[1] · Francisco A. Tomás-Barberán[3] ·
Christine Morand[8] · Rodrigo Feliciano[18] · Rocío García-Villalba[3] · Mar Garcia-Aloy[5,6] · Pedro Mena[7]

Abstract

Purpose The quality of the study design and data reporting in human trials dealing with the inter-individual variability in response to the consumption of plant bioactives is, in general, low. There is a lack of recommendations supporting the scientific community on this topic. This study aimed at developing a quality index to assist the assessment of the reporting quality of intervention trials addressing the inter-individual variability in response to plant bioactive consumption. Recommendations for better designing and reporting studies were discussed.

Methods The selection of the parameters used for the development of the quality index was carried out in agreement with the scientific community through a survey. Parameters were defined, grouped into categories, and scored for different quality levels. The applicability of the scoring system was tested in terms of consistency and effort, and its validity was assessed by comparison with a simultaneous evaluation by experts' criteria.

Results The "POSITIVe quality index" included 11 reporting criteria grouped into four categories (Statistics, Reporting, Data presentation, and Individual data availability). It was supported by detailed definitions and guidance for their scoring. The quality index score was tested, and the index demonstrated to be valid, reliable, and responsive.

Conclusions The evaluation of the reporting quality of studies addressing inter-individual variability in response to plant bioactives highlighted the aspects requiring major improvements. Specific tools and recommendations favoring a complete and transparent reporting on inter-individual variability have been provided to support the scientific community on this field.

Keywords Clinical trials · Reporting quality · Inter-individual variation · Quality index · Plant bioactive ·
Recommendations · Guidelines

✉ Aleksandra Konic Ristic
  aleksandra.konicristic@ucd.ie

✉ Pedro Mena
  pedromiguel.menaparreno@unipr.it

Extended author information available on the last page of the article

Introduction

A large body of evidence supports the notion that bioactive compounds present in plant foods [e.g., (poly)phenols, carotenoids, glucosinolates, etc.] have numerous beneficial effects on human health, underlying the association between the habitual consumption of plant-based diets and the reduced risk of age-related chronic diseases [1, 2]. However, data from clinical trials aiming to establish the cause–effect relationship are often inconclusive and even contradictory [3, 4]. It has been suggested that this may be, at least partly, the result of the differences in the bioavailability of plant bioactives among individuals, as well as the variation in

their effects on specific functional biomarkers, either physiological or biomarkers of risk [5]. A clear understanding of all the factors responsible for the inter-individual variability (IIV) in response to plant bioactives is the vital part of knowledge that is still lacking, yet it is considered crucial for the ultimate positioning of plant bioactives on the roadmap to optimal health. To gain this seminal knowledge, the complex interactions of plant bioactives with the factors that drive the genotype and phenotype of individuals should be assessed [5–7].

A recent series of systematic reviews and meta-analyses of human studies addressing the IIV in response to different plant bioactives concluded that the number of trials reporting between-subject variations is, in general, very low [8–10]. At the same time, since most studies are not initially designed to address IIV, they are often underpowered within groups and unbalanced between them, with multiple sources of bias including selective reporting, insufficient description of subjects' characteristics and inaccurate reporting on statistics, as well as providing inadequate conclusions on the observed effects. Consequently, although in most of the cases the quality of protocol design and reporting of the initial study respond sufficiently to general recommendations and requirements [11–13] studies were often considered as non-adequate in terms of post hoc subgroup statistics, regressions, or other similar approaches in the analysis. It is thus evident that the design and reporting of human trials addressing the between-subject variation are crucial and should be improved. Moreover, a better way of designing and reporting human intervention trials on the effect of plant bioactives while considering individual differences would help to further summarize and analyze aggregated data through systematic reviews and meta-analyses. An inadequate reporting quality of studies often makes them non-eligible to be included in meta-analyses or it may introduce a significant bias in the analysis, compromising the accuracy and reliability of the conclusion [10, 14]. Nevertheless, despite the multiple benefits related to appropriate design and reporting of studies dealing with IIV in response to plant bioactive consumption, there is a lack of suitable recommendations and/or guidelines supporting the scientific community and favoring improvements in the way data are assessed and disseminated. Unfortunately, the quality assessment tools that are available to date [11, 15–17] do not fit sufficiently the scientific questions and purposes of this specific type of reporting of intervention trials on food products or supplements rich in plant bioactives.

This study aimed at proposing a specific quality index (QI) as a tool to be used in the assessment of the reporting quality of human intervention trials addressing the IIV in response to plant bioactive consumption (so-called the "POSITIVe quality index"). The IIV QI was tested for its reliability, validity and responsiveness. The steps carried out to develop this QI have also been thoroughly reported as a roadmap and recommendations for its application to better design and report human intervention trials in the field have been provided.

Methods

Rationalizing the need to develop an additional tool to assess reporting quality and expert agreement

This study was performed as part of the COST Action FA1403 POSITIVe (https://www6.inra.fr/cost-positive), a collaborative and multidisciplinary pan-European network of more than 300 researchers. One of the main objectives of the Action was to identify the main factors responsible for the observed IIV in response to plant food bioactives intake, based on available scientific evidence and generated new knowledge. While dealing with this objective in the early phases of the Action, it was noted that the reporting quality of human studies in the field was often limited and inadequate. This issue was further discussed by ten experts of the Action Think Tank Group (addressed in the following text as "score developers"). One of the main reasons hypothesized was the lack of generally accepted and routinely applied recommendations to report between-subject variation, either as primary or post hoc analysis. The strategy to solve this gap was defined and it consisted of several initial steps: (1) to review the existing literature and identify previous relevant guidance and assessment tools; (2) to seek relevant evidence on the quality of reporting in published research articles; (3) to identify key information related to the potential sources of bias in such studies; and (4) to use the POSITIVe Action network for expert opinions and resources.

The literature search was performed through the Equator Network Library resources (http://www.equator-network.org/library/), Medline, Embase and Cochrane databases, and through a Pubmed search using specified search terms (e.g., "quality assessment", "tools", "reporting", "guideline(s)", "inter-individual", "quality index", "recommendation").

The retrieved and critically selected literature included: (1) tools often used for assessing reporting quality of individual studies [18] or tools for assessing reporting bias during systematic reviews and meta-analyses [19]; (2) a scarce number of available resources that specifically address IIV in clinical trials [20–23]; (3) papers retrieved during the ongoing systematic searches and meta-analyses on IIV in response to plant bioactives; (4) two position papers of the Action [5, 6]; (5) several papers on "controversies" on clinical trial methodology and reporting [24–26]; and (6) available strategies on creating and developing tools and guidelines [27–29].

Based on the information from the extracted literature, the "score developers" made a list with up to 23 criteria addressing all relevant parameters that may be considered as specific for data reporting in studies assessing IIV in response to plant bioactives intake (Fig. 1).

In the next step, the "score developers" designed a questionnaire entitled "How to Report Inter-Individual Variability in Publications" that was sent to 311 members of the POSITIVe network. It included questions addressing expert opinion on (i) the need for a specific assessment tool to be developed; (ii) the familiarity with the Jadad score used to assess the quality of reports of randomized trials; (iii) the need for extension for this specific purpose; (iv) the interest of reviewers and journal editors in the network to adopt and implement a quality index score (QIS) in the reviewing process; and, as a crucial part, (v) members were asked to select a number of criteria they consider relevant enough to be included in the score, from the list of 23 criteria made by the "score developers".

Selection of categories and parameters to be addressed in the assessment and development of the IIV quality index scoring system

After collection and evaluation of the questionnaires, "score developers" defined the categories and selected the parameters within each category to be included in the score. Parameters selected by 50% or more experts were directly taken into consideration to be included in the IIV QI. Parameters that were considered important by 40–50% of experts were additionally discussed and evaluated by score developers, while parameters selected by less than 40% of participants were excluded from the score. As a result, a list of 11 parameters grouped in 4 categories was defined for the design of the QI and the development of recommendations/guidelines for data reporting on IIV (Table 1).

The next step was the creation of the first version of a dictionary, with detailed definitions of the conditions related to data reporting for each parameter and with assigned marks, as a base for the calculation of the QIS. Marks were assigned to each parameter and its related conditions while reporting as follows: if a particular parameter is not considered in the study at all, its score is 0; if it is reported but not informative enough to completely describe IIV, its score is 0.5; and, if it is considered and completely illustrate IIV, its score is 1. The exception to this scale was the last category, related to the individual data availability. Considering the importance of this parameter for the accurate assessment of IIV, developers modified the scale to 0–1–2, to increase the weight of this parameter in the score. The dictionary of the QI and parameter marks are presented and fully described in the results section.
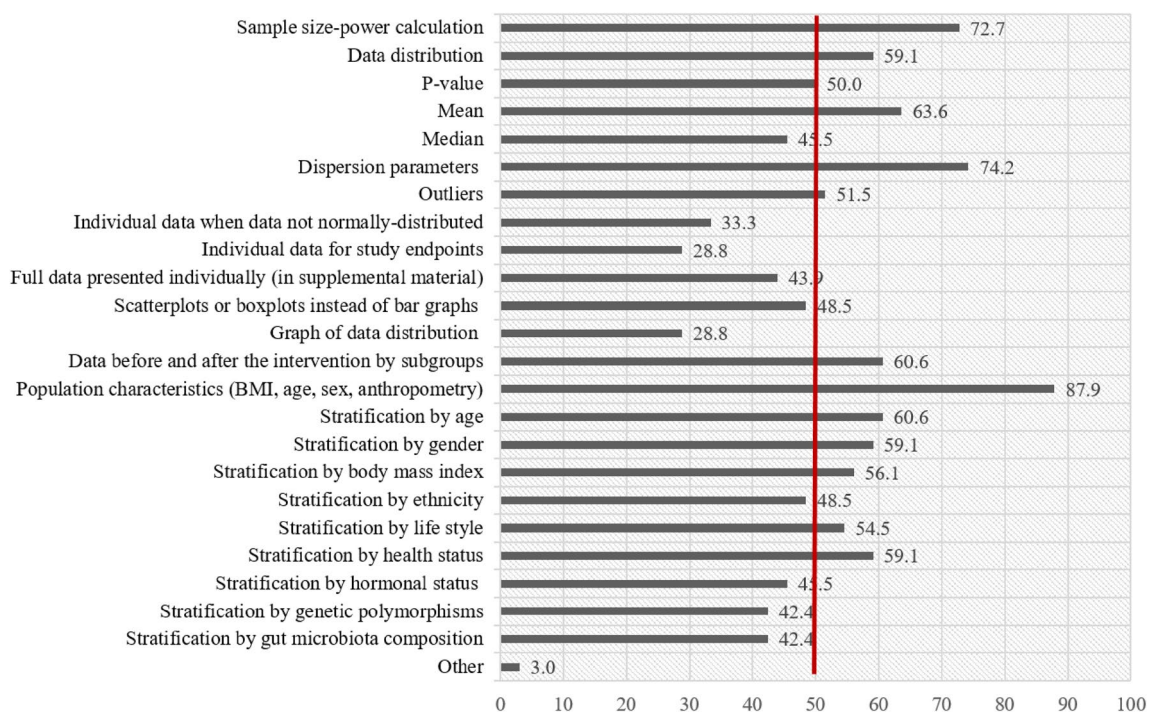

Fig. 1 Percentage of experts who considered the listed parameters important to be reported when assessing inter-individual variability in response to consumption of plant food bioactives

Table 1 Dictionary of conditions related to data reporting on IIV and associated scores for the evaluation of quality of data reporting in intervention studies dealing with plant bioactives

Category	Parameters	Condition	Score
Statistics	Sample size-power calculation	Authors reported on: (1) power calculation focused on assessing inter-individual variation based on primary OR the most limiting outcome AND (2) on all the data used in power calculation (% of statistical power, significance level, expected dropout rate, expected difference between groups of the mean or % with the event) AND (3) the resulting sample size per each group	1
		Authors did not describe sample size taking into account all the three previous conditions	0
	Data distribution	Authors specified the test used for normality (OR indicate something related to data normality, for instance, log transformation)	1
		Authors did not report any information related to the normality or distribution of the data in general	0
	p value	Authors reported p value that support IIV (e.g., p value related to the examination of differences between two or more factors affecting IIV such as sex, age, genotypes, etc.)	1
		Authors did not report any p values related to the IIV	0
	Effect size	Authors reported the magnitude of the IIV for the selected outcome(s) by standardized mean differences as an index of effect size (Cohen's d, % of coefficient of variation, etc.) or any parameters related to the effect size suitable for the conducted statistical tests	1
		Authors did not indicate any parameters related to the effect size i.e., magnitude of the IIV for the selected outcome(s) (any of standardized mean differences was not reported)	0
Reporting	General characteristics of the subgroups where IIV was evaluated	Authors reported on one or more general characteristics (for instance, ethnicity, BMI, age, gender, smoking status, etc.) for each of the subgroups where IIV was evaluated	1
		Authors did not report any of general characteristics for the study sample (for instance, ethnicity, BMI, age, gender, smoking status, etc.) on each subgroup where IIV was evaluated	0
	Data reporting for end-points by subgroups	Both pre- and post-intervention data (or post-intervention data as % change with respect to a provided baseline value) were reported for different subgroups where IIV was evaluated	1
		Post-intervention data (or % change without a provided baseline value) are provided by each subgroup where IIV was evaluated	0.5
		Neither pre- nor post-intervention data were reported for different subgroups where IIV was evaluated	0
	Measures of central tendencies and dispersion parameters	Authors reported on one or more measures of central tendencies (mean, median, etc.) AND one or more dispersion parameters (standard deviations, standard error, interquartile range, 95% confidence interval, etc.) for EACH subgroup where IIV is evaluated	1
		Authors did not report any measures of central tendencies or dispersion measures for subgroups where IIV was evaluated, regardless of reporting these parameters for the total sample	0
	Outliers	Authors indicated outliers AND described them (explained the reason for treating them as outliers)	1
		Authors indicated that outliers existed and that were excluded from the analysis but without describing them	0.5
		Authors did not indicate any information related to outliers	0
Data presentation	Tables	Tables contain additional measures of variability (min–max, interquartile range, outliers values, etc.) or individual measures (responders/non-responders, etc.)	1
		Tables did not contain any extra measures of variability (min–max, interquartile range, outliers values, etc.) nor individual measures (responders/non-responders, etc.)	0
	Graphs	Authors presented data for the primary outcome by scatterplots, boxplots or heat maps	1
		Authors presented data by histograms for a primary outcome OR as scatterplots and boxplots for secondary outcomes	0.5
		Data are graphically presented as bar chart, curves, etc. (for any of the study point-before or after the intervention) but not as scatterplots, boxplots, heat maps or histograms	0
Individual data availability	Presentation of full data and population characteristics	Authors provided individual data for each end-point, together with the characteristics of the samples on the individual level, in the paper or in the supplemental material	2
		Authors provided individual data at each end-point (even presented in the figures) but without any additional characteristics of the sample on the individual level	1
		Authors did not provide individual data	0

Testing the IIV quality index score validity, reliability, and responsiveness

In the next phase, the QI was tested and validated by evaluating and scoring existing studies, on the basis of their comprehensiveness of data reporting on IIV. The evaluation and scoring were performed by nineteen experts ("evaluators"), all members of the Think Tank Group of the Cost Action POSITIVe. Evaluators had previous experience in clinical trials on plant bioactives, as shown by their publication records, and they were all involved in the ongoing systematic reviews and meta-analyses performed as part of the objectives of the Action [5, 8–10]. The evaluators screened the dossiers of peer-reviewed articles retrieved within the systematic searches they were involved in and identified 30 articles that specifically reported IIV in response to plant bioactives consumption. The total number of 30 studies was considered sufficient for the purpose of testing the IIV QI.

In the first step, the comprehensiveness of the parameters and the dictionary was tested in a pilot trial by evaluating five studies. These studies were selected from the comprehensive list of 30 studies, using computer-assisted random selection from 2 subgroups, resulting in 2 random studies on bioavailability and 3 on the effects of plant bioactives. Each study was evaluated independently by two or three different evaluators, and the final scoring was done based on their consensus. After the pilot testing phase, all evaluators provided their critical opinion to the developers and helped in revising the dictionary to be clear and user friendly. Revisions were related to the definitions of conditions and regrouping parameters between categories. Once the final version of the dictionary for the QI was created, 30 studies, (including 5 using in the test phase) were evaluated and scored in the same way as in the testing phase [30–59].

After the evaluation, data reporting quality of each study was assessed by the total score and four sub-scores related to the four categories considered: Statistics, Reporting, Data presentation and Individual data availability (Table 1). In addition to the evaluation carried out by employing the IIV QIS, evaluators assessed the overall quality of data reporting on IIV for each study using a qualitative scale and characterizing them as weak, mild, or good based on their personal opinions as experts. QI was then validated by comparing these two methods of evaluation.

Statistics

Accuracy of QI was examined by analyzing relations between the quality of the studies as assessed by experts using quantitative scale (weak, mild, and good) and the QIS. The ordinal variable was created for the quality level assessed by experts (weak = 1, mild = 2, good = 3). The overall quality score was calculated for each study as the sum of all marks given to each study divided by 11 (number of selected parameters). Completeness of reporting within each defined category was calculated for each study as the sum of all marks assigned to the parameters from the category divided by the number of parameters for that particular category. In this way, completeness of reporting was standardized for all categories. Spearman's correlation coefficient was calculated for the relation between overall IIV QIS and the quality level assessed by experts. Cohen's kappa coefficient was calculated to test the agreement between tertiles of the overall IIV QIS and quality levels assessed by the experts. Impact of completeness of each defined category within the QI on the quality levels assessed by experts was also assessed by the Spearman correlation coefficient. A value of $p < 0.05$ was taken to indicate a significant result. All analyses were performed using SPSS statistical software (IBM SPSS statistics 20, SPSS Inc., Chicago, IL USA).

Results

Questionnaire results

The questionnaire 'How to Report Inter-Individual Variability in Publications' was answered by 66 experts involved in the POSITIVe network, resulting in a response rate of 21%. The majority of responders (96%) considered the development of the QI important for the assessment of the quality of data reporting on IIV in publications dealing with plant bioactive consumption. In general, experts had different approaches during the selection of parameters from the list. The selection of fewer than ten parameters (more critical approach) was observed in 44% ($n = 29$) of those who answered the questionnaire. This subgroup of experts was focused on statistical parameters (sample size, normality, p value related to IIV), parameters related to the measures of central tendencies and dispersion, and parameters related to the population characteristics and stratification by different factors that could affect IIV. The rest of the experts were prone to choose more than 10 parameters, 24% of them selected between 11 and 15 parameters, while 32% of them considered more than 15 parameters as important. Answers provided by all responders were taken into consideration in further developing the QI. Percentage of experts, who considered the listed parameters important to be reported when assessing IIV in response to plant bioactives, is presented in Fig. 1. Sample size calculation, dispersion parameters, and population characteristics were selected by more than 70% of experts, as the most important parameters related to IIV. Data distribution, p value, mean, outliers, data before and after intervention by subgroups, and stratification by different factors (age, gender, body mass index, lifestyle, and health status) were considered important by more than 50%

of experts. However, stratification by ethnicity, hormonal status, polymorphism, and gut microbiota composition were selected by 40–50% of experts, as well as median, full data presented on individual level provided in the supplementary material, and data depicted in scatter or box plots. Graphical presentation of data distribution, individual presentation of non-normally distributed data, and individual data for study end-point were assessed as important by less than 35% of experts. Additional parameters under the option "other" were suggested by 12 experts. Only one of them (coefficient of variation) was related to data reporting from the statistical point of view, while all other parameters were more general and not directly related to the data reporting but factors important for IIV, like dietary habits, physical activity, etc. The experts were asked about the Jadad scale for reporting randomized trials, and 36.4% of them stated that they were familiar with this scale developed in 1996 by Jadad et al. [15] while 59% of responders said that it would be important to supplement the Jadad scale with the QI related to IIV.

Among the experts who answered the questionnaire, there were 16 (24%) members of editorial boards in peer-reviewed scientific journals, with the SJR ranking for these journals being between 0.22 and 1.53 in 2018, including 7 with the SJR above 1. The majority of experts (80%) claimed that journal editors might be interested in the QI to be used during the evaluation process of manuscripts dealing with IIV in response to plant bioactives.

Development of the quality index and the dictionary as associated explanatory document

The final version of the dictionary with defined conditions for each parameter and assigned scores is presented in Table 1. Sample size/power calculation, distribution of the data (normality), and p value related to the IIV were grouped in one category (*Statistics*). Additionally, after testing the first version of the dictionary for the QIS, evaluators stressed the importance of reporting on the effect size of the applied statistical tests, as an important parameter for the complete understanding of the p value significance [60]. Finally, the sum of scores based on the first category (*Statistics*) reflects on the quality of data reporting concerning four parameters: sample size/power calculation, data distribution, p value, and effect size. All parameters in this category could be assessed by a dichotomous score of 0 or 1.

Another set of six parameters listed in the questionnaire were regrouped into four and integrated into the second category (*Reporting*) (Table 1). Reporting on general characteristics of the subgroups where IIV is generally evaluated (age, gender, body mass index, smoking, etc.) was the parameter most often selected by experts (87.9%). Since they had diverse opinions about stratification according to different

characteristics, it was decided to keep this parameter open for any characteristic collected for the possible study subgroups. Reporting on data for study end-points by subgroup (before and after the intervention) was also included within the *Reporting* category. Furthermore, measures of central tendencies and dispersion parameters reported for each subgroup were merged and included in this category as a single parameter. Reporting on outliers was integrated into the *Reporting* category as the fourth parameter. Data reporting on end-points by subgroups and on outliers could be scored by values 0, 0.5, or 1, depending on the comprehensiveness of the reporting (Table 1). On the other hand, reporting on general characteristics and measures of central tendencies and dispersion parameters could be scored by 0 or 1. In conclusion, overall score related to the *Reporting* category reflects the quality of reporting on four parameters: general characteristics of the subgroups, data reporting for end-points by subgroups, measures of central tendencies and dispersion parameters, and outliers.

The third category considered important by developers, taking into account both tables and graphs (Table 1), was *Data presentation*. Though the complexity of tables was not listed as a parameter in the questionnaire, after testing, evaluators suggested to include this parameter. Its purpose is to assess reporting on additional values that could reflect IIV (min, max, interquartile range, number of responders/ producers, etc.), apart from common measures of central tendencies and dispersion parameters. Presenting data as scatter plots or box plots instead of bar charts was considered important by 49% of experts. Developers decided to extend this parameter to assess the global quality of graphical data presentation, making a distinction between presenting data on primary and secondary outcomes. It means that, in the final version of the dictionary, the scatter plots, box plots, or heat maps that depict data related to the primary outcome are considered as the most useful (score = 1); histograms depicting primary outcome data or scatter plots/box plots/heat maps depicting secondary outcomes are considered as not so informative but still useful ways of illustrating IIV (score = 0.5); and bar charts, curves, etc. for any outcome are considered as not helpful in the assessment of IIV (score = 0).

The fourth category was related to the availability of individual data, i.e., transparency of analyzed data set and the possibility of further analysis. Developers distinguished three different levels within this category. Individual data available for each end-point reported together with the characteristics of the study participants on an individual level are the most appreciated option (score = 2). Individual data reported for each end-point but without any additional characteristics of study participants are still considered useful but less than the previous option (score = 1), while studies

that did not show any individual data are not scored for this category (i.e., score = 0).

Validation of the quality index-evaluation of collected studies

QIS and categories sub-scores were calculated for 30 studies (Supplementary Table 1). Quality of data reporting on IIV for these studies was additionally evaluated by experts using qualitative scale, and 2 studies were assessed as weak, 12 studies as mild, and 16 studies as good. A weak agreement was found between terciles of overall QIS and the three levels of quality (weak, mild, and good) assessed by experts (Cohen's $k = 0.216$, $p = 0.054$). The significant agreement between both methods was confirmed by Spearman's rank correlation coefficient ($r = 0.697$, $p < 0.001$).

Numbers and percentages of studies that reported any data on the selected parameters (scored either 1 or 0.5 according to the dictionary) are presented in Table 2. Significant correlations were found between the completeness of particular score categories (*Statistics* and *Reporting*) and quality levels assessed by experts (Spearman's $r = 0.519$, $p = 0.003$; $r = 0.509$, $p = 0.004$, respectively). On the contrary, *Data presentation* category was inversely correlated with the expert's opinions ($r = -0.365$, $p = 0.047$). Individual data availability, as an independent parameter, were not analyzed in this way since only four studies provided data, but without additional characteristics of the groups where IIV was evaluated. Weak, mild, and good studies that reported on selected parameters are presented in Fig. 2. *Statistics* category was found as the most important for high-quality data reporting on IIV. All studies assessed as good by experts reported on, at least one, parameter from the *Statistics* category, while 75% of them reported on two or more parameters from this category.

Moreover, 91% of those that reported on effect size were assessed as good by expert's independent opinion. Only one study reported on sample size as expected (i.e., as described in the dictionary). Studies that reported on data distribution and p value related to IIV were assessed as good in 67% and 61% of cases, respectively (Fig. 2). Regarding the *Reporting* category, most of the studies assessed as good by experts took into consideration parameters related to this category (Fig. 2). Results showed that 63% out of 21 studies that reported on data for end-points by subgroups were categorized as good by evaluators, and there were no studies characterized as weak. These results were similar for studies reporting on measures of central tendencies and dispersion parameters by subgroups (68% of them characterized as good). Twelve out of 20 studies (60%) that reported on general characteristics of the sample, where IIV was evaluated, were characterized as good. Of note, studies reporting outliers were classified as mild according to experts'

opinion. Last, as suggested by the inverse correlation aforementioned, graphs and tables as defined in the dictionary (*Data presentation* category) were not key parameters for a comprehensive explanation of IIV. Eight out of 14 studies (57%) that presented graphs as defined in the dictionary were assessed as weak or mild. In the case of the tables, only two studies reported on additional measures of variability (min–max, interquartile range, etc.) or individual measures (responders/non-responders, producers/non-producers, etc.), and they were assessed as mild.

Discussion

The main aim of this work was to develop a tool to support the assessment of the reporting quality of individual studies addressing either predefined or post hoc IIV in response to the consumption of plant bioactives. The developed tool, the "POSITIVe quality index" comprised 11 individual parameters classified into 4 categories and weighed/scored by criteria described in the accompanying explanatory material—the dictionary—(available at Supplementary material 2). The whole development process was performed as part of the activities of the COST Action POSITIVe and by the stepwise processes proposed by the Equator network collaborators for developers of health research reporting guidelines [27]. Other available and relevant recommendations on designing tools and general guidelines in the area of medical research, reporting on research data but also the dissemination of outcomes and medical practice [28, 29], were also taken into account.

POSITIVe quality index parameters and recommendations

All the parameters included in the POSITIVe QI demonstrated to be useful for an adequate reporting of the IIV associated with the consumption of plant bioactives, both in bioavailability and bioactivity studies. Instead of emphasizing the parameters thoroughly described in the dictionary (Table 1), some aspects are worth discussing. During the selection of individual parameters to be included in the POSITIVe QI, individual data availability was considered a useful parameter to be reported to understand IIV fully. Moreover, as defined in the dictionary, individual data for each end-point together with the individual subject characteristics (age, gender, ethnicity, etc.) are of the greatest value, not only for understanding the IIV at a trial level but also for further data aggregation and meta-analyses [61]. Although the practice of data sharing significantly increases in the clinical research community, we found only 4 out of 30 studies on plant bioactives that provided individual data (but without additional characteristics of the subjects at

Table 2 Reporting on parameters within defined categories [n (%)] and completeness of reporting (%) within each category by quality of studies assessed by experts' opinion

Category	Parameter	Weak ($n=2$)		Mild ($n=12$)		Good ($n=16$)		Total ($n=30$)		Completeness vs. quality levels assessed by experts	
		n (%)	Completeness	n (%)	Completeness	n (%)	Completeness	n (%)	Completeness	Spearman's correlation coeff.	p value
Statistics	Sample size-power calculation	0 (0%)	37.5%	0 (0%)	20.8%	1 (6%)	48.4%	1 (3%)	36.7%	0.519	0.003
	Normality/distribution of data	1 (50%)		2 (17%)		6 (38%)		9 (30%)			
	p value related to IIV	1 (50%)		8 (67%)		14 (88%)		23 (77%)			
	Effect size	1 (50%)		0 (0%)		10 (63%)		11 (37%)			
Reporting	General characteristics of the subgroups where IIV was evaluated	1 (50%)	25%	7 (58%)	44.8%	12 (75%)	59.4%	20 (67%)	51.3%	0.509	0.004
	Data reporting for end-points by subgroups	0 (0%)		8 (67%)		13 (81%)		21 (70%)			
	Measures of central tendencies and dispersion parameters	1 (50%)		6 (50%)		14 (88%)		21 (70%)			
	Outliers	0 (0%)		2 (17%)		0 (0%)		2 (7%)			
Data presentation	Tables	0 (0%)	25%	2 (17%)	33.3%	0 (0%)	14.1%	2 (7%)	22.5%	−0.365	0.047
	Graphs	1 (50%)		7 (58%)		6 (38%)		14 (47%)			
Individual data availability		0 (0%)		2 (17%)		2 (13%)		4 (13%)		–	–
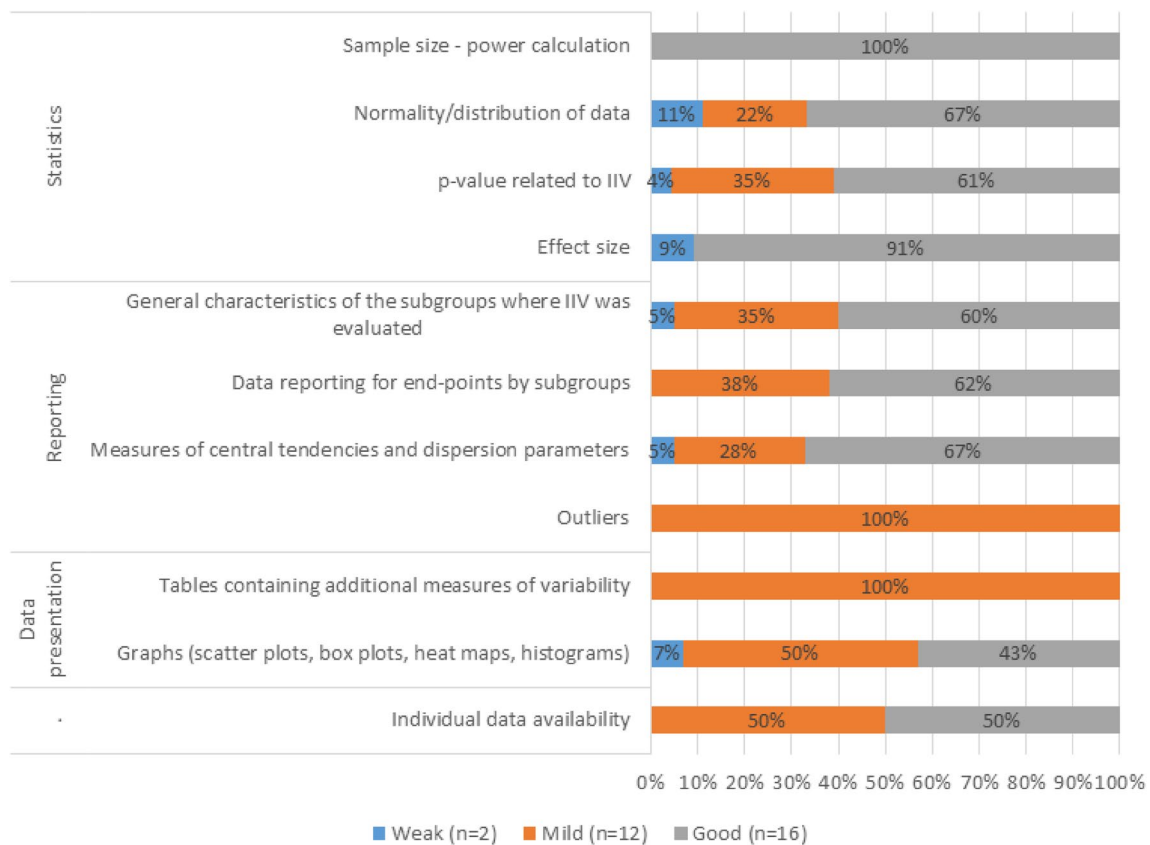
Fig. 2 Number and percentage of good, mild, and weak studies with respect to the reported parameters within defined categories

the individual level) [30, 31, 52, 55]. Therefore, in order to contribute to better handling and understanding the IIV in human intervention trials, it is highly recommended for authors to prepare their data for sharing either in publications or in one of the existing research data repositories such as ClinicalTrials.gov or repositories of the Open Research Data Pilot of the European Commission, based on an adequate Data Management Plan [62]. Ohmann et al. summarized all the principles and recommendations for Individual Participant Data (IPD) sharing that should be followed in data sharing processes [63]. In case that authors decide not to share IPD, comprehensive data reporting on other categories of the QI are recommended all along the scientific process.

Regarding the study design procedure, proper *Statistics* should be taken into account, starting from the sample size calculation. An adequate calculation of the sample size is an essential element of high-quality data reporting on IIV. After the evaluation of the studies regarding power calculation and sample size, we found that conditions given in the dictionary are too demanding for this research area. There are still not enough studies, dealing with the effects of plant bioactives that reported on IIV between different groups, that authors could use to learn

the population standard deviation of particular groups and related interventions. That is likely the reason why we found only one study [31] that took into consideration IIV for the sample size calculation. However, it is highly advisable to look for all studies that reported on similar results and, if they exist, to take into account reported IIV to calculate sample size.

Reporting on the distribution of data when dealing with IIV is as important as for all other data reporting cases, but we want to emphasize the importance of checking data distribution and other assumptions that should be met to get accurate statistical results. Misunderstandings of assumptions that need to be satisfied before employing parametric tests happen often. For example, the assumption for dependent *t* test that the sampling distribution of the differences between scores of two measures should be normal is usually misinterpreted by a normal distribution of scores themselves. The assumption of normal distribution within the groups, which is required when employing one-way ANOVA, is misinterpreted as the normal distribution of the total sample, etc. Assumptions for the extended list of parametric tests are explained in detail by Field et al. [64]. Visual methods for checking normality like histograms, box plots, stem-and-leaf plots, etc. could be helpful for large data sets, since statistical

tests (e.g., Kolmogorov–Smirnov test, Shapiro–Wilk test, etc.) could be significant, i.e., rejecting the hypothesis of normal distribution even if deviations from normality are small. On the other hand, for small samples (< 30), statistical tests are fully necessary [65]. Shapiro–Wilk test is recommended as the best choice for testing the normality of data [66]. Then, based on data distribution, the proper statistics should be used and justified, since, for example, the presence of responders and non-responders (or producers and non-producers) may condition the data distribution. Although it could be assumed that researchers are checking the assumptions needed to use a particular test, reporting on that would indicate undeniably that those assumptions have been assessed.

Reporting on *p* value, without reporting on central tendencies and dispersion measures or the change by subgroups, is not as informative as *p* value reported together with these parameters. For example, reporting on *p* value related to different effects between men and women, as an explanation of a scatter plot, without reporting on mean and SD for each subgroup, is not as informative as it would be if authors provide these data numerically, especially if differences are small. This type of reporting cannot be used in further meta-analysis related to the IIV, unnecessarily limiting the understanding of IIV by not reporting on data that definitely exist. Moreover, these data could not be helpful for sample size calculation in future studies. Thus, reporting on *p* value should always be followed by reporting on central tendencies and dispersion measures for compared groups.

An additional parameter important for understanding the statistical significance of the effects of the intervention (*p* value) is the effect size. Even though the *p* value provides the information that effect of the intervention exists or not, it does not indicate the size of the effect (magnitude of the difference between groups/treatments) [67]. As shown in the results section, 91% of studies that reported on effect size were assessed by experts as good, regarding the quality of data reporting on IIV. Thus, it is highly recommended to report on the effect size together with the *p* value. In the studies evaluated, an informative way to report the effect size was using percent coefficient of variations, but further guidance on how to calculate and interpret an effect size for different types of analysis is summarized by Durlak et al. [68].

Although 61% of studies that presented data graphically as defined in the dictionary were assessed by experts as weak or mild, we still recommend proper graphical representation of data. As the effect size provides additional explanation to the *p* value, appropriate graphs could disclose statistics reported in tables. This is especially important for small sample sizes, as it is usually the case of nutritional intervention studies. Among the different graphs that can be used to represent IIV, scatterplots of raw data could be the most useful graphs regarding transparency of the results when dealing with small sample sizes. Such graphs could clearly show where standard deviations come from, particularly if subgroup characteristics are reported. Box plots are also a very useful way of data presentation since it shows outliers and variation.

Nevertheless, as box plot summarizes data, they are more meaningful for large sample sizes. The same applies to histograms, they are considered useful in depicting the distribution of large samples, but they are hard to understand and interpret regarding the IIV for the small samples [65, 69]. Bar charts are not recommended since they cannot say much more than a table and, moreover, they do not help reveal the distribution of data at all, since the same bar chart could be created based on different distributed data sets [69].

Another way of favorable data reporting are tables that consider criteria beyond the central tendencies and dispersion measures, including more parameters that could uncover data distribution like min–max, median, interquartile range, coefficient of variation, etc. In this way, readers get a clearer picture of data distribution and direction of variation. For instance, Brindani et al. reported on central tendencies and described the data distribution from their sample by additional parameters [70]. This type of approach may be useful to better highlight the IIV.

In the end, reporting on outliers could be considered a good way to present subjects responding in an atypical way and, thus, indicate eventual IIV. The search for homogenous data has somehow demonized outliers in research. However, when dealing with IIV, outliers can be a precious resource to better understand the differential response to the consumption of plant food bioactives and could serve to further explore the reasons behind extreme responses. Once confirmed that any potential outlier is not the result of an analytical artifact, it should be considered as robust data indicating individual variability. Their exclusion from statistical tests may be justified but reporting on them is advisable.

Overcoming challenges in reporting inter-individual variation

The quality of reporting of clinical trials is a critical part of their overall quality as it allows readers to judge other elements in the quality domain (the design, conduction, analysis and clinical implications) and make conclusions about the reliability of their reported benefits and harms [71]. For more than two decades, an enormous effort was put by experts in clinical research into increasing the quality of reporting, with the introduction of the CONsolidated Standards Of Reporting Trials (CONSORT) Statement being a pillar. The initial statement was published in 1996 [72], further updated in 2001 [11], and 2010 [73], and supplemented with 15 official guidelines for different types of studies and 14 official extensions that address reporting of

different aspects such are study designs, interventions or type of data (http://www.equator-network.org/). It has been confirmed that the introduction of these guidelines and their acknowledgment by the scientific community significantly improved the reporting quality.

There are still areas in trial reporting that remain insufficiently described and defined, with contradictory opinions or addressed even as "controversies" or (mis)uses in clinical research, such as multiplicity of data, co-variate adjustment vs. subgroup analysis, assessing individuals benefits and risks from clinical trials data, or interpretation of surprising results [24]. At the same time, most of these challenging topics are considered highly relevant for assessing the impact of IIV, using trial data for identifying those who will benefit the most, or with the least harm. For example, it is widely accepted that subgroup analyses, especially if performed post hoc, may be misleading and could bring a high risk of false conclusions that very often cannot be confirmed by subsequent studies and may have detrimental consequences [13]. However, if conducted and reported properly, analyses of IIV on trial level (e.g., effects in different subgroups) could lead to more precise recommendations, provide supporting evidence for making substantiated decisions, and could help (re)building the trust of patients/consumers [74]. Acknowledging this, the International Committee of Medical Journal Editors recommends stratifying reporting stating: "Separate reporting of data by demographic variables, such as age and sex, facilitate pooling of data for subgroups across studies and should be routine, unless there are compelling reasons not to stratify reporting, which should be explained" [75]. A list of additional sets of criteria to test the credibility of subgroup analyses includes testing whether the effect (1) is consistent across studies, (2) can it be the result of a chance, (3) is there a biological reason for the observed effect, (4) is the reported evidence from within- or between-study comparison, (5) is it defined a priory or it is a post hoc [76]. More detailed guidelines on statistical issues associated with subgroup analysis have been addressed in previous publications [73, 77] with the general conclusion that more clear and complete reporting of subgroup analyses and similar models of analyzing IIV should be encouraged. Last, efforts should also be paid in reducing, as much as possible, sources of variability not directly related to physiological conditions but to analytical constraints. In this sense, errors in the selection and measurement of some biomarkers of intake and effect that may condition the assessment of the individual response should be overcome. A deep effort and investment should be carried out in the development of standardized methodologies for the analysis of specific, reliable, and reproducible biomarkers [78, 79]. This would further favor the comparison of the reported results.

Authors' awareness is likely another challenge to be overcome. Although the interest in the differential individual response to the consumption of plant food bioactives is growing, as demonstrated by the increasing number of publications in the field [15], the number of works dealing with this topic is rather scarce if compared to the number of publications in the field. Besides more research approaches addressing this topic from the study design, further efforts are required to report the putative IIV existing in any study. By considering the recommendations on reporting indicated in the dictionary of this QI (Table 1), authors may endow their works with valuable information on the existing IIV. Small efforts during the preparation of their manuscripts may provide a plethora of valuable information and benefits. The authors would benefit from the increased quality of their manuscripts and research, while the whole scientific community would benefit from the availability of this key information. This work presents the consensus of a significant part of the scientific community in the field. Adopting this consensus on reporting guidelines will enable the full appraisal of the trial conducted and will boost the possibilities to pool research data to gain further evidence. In summary, minor changes in the way we report data may lead to major developments in the field of plant bioactives at the individual level, moving with the times to favor suitable strategies for personalized nutrition.

Finally, although the POSITIVe QI was tailored to address the need for a better reporting of inter-individual variation in response to plant bioactives, it might be considered for non-plant bioactive compounds and as a starting point to address similar issues in other related areas, including precision medicine, public health, pharmacokinetics or toxicology [80–82]. Further validation of its use on these fields by collaborative networks is suggested.

Authors' contributions MN performed data analyses, was involved in the interpretation of results, and drafted the manuscript; AKR contributed to the design and the protocol of the study, performed data analyses, was involved in the interpretation of results, and drafted the manuscript; AGS, GI, RF, RGV, and MGA contributed to the design and the protocol of the study and were involved in the interpretation of results; PM contributed to the design and the protocol of the study, was involved in the interpretation of results, critically reviewed the manuscript, and had primary responsibility for final content. All authors conducted a literature review, read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

1. Fraga CG, Croft KD, Kennedy DO, Tomás-Barberán FA (2019) The effects of polyphenols and other bioactives on human health. Food Funct 10:514–528. https://doi.org/10.1039/c8fo01997e

2. Aune D, Giovannucci E, Boffetta P, Fadnes LT, Keum N, Norat T, Greenwood DC, Riboli E, Vatten LJ, Tonstad S (2017) Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality—a systematic review and dose-response meta-analysis of prospective studies. Int J Epidemiol 46:1029–1056. https://doi.org/10.1093/ije/dyw319

3. Cicero AFG, Fogacci F, Colletti A (2017) Food and plant bioactives for reducing cardiometabolic disease risk: an evidence-based approach. Food Funct 8:2076–2088. https://doi.org/10.1039/c7fo00178a

4. García-Conesa MT (2017) Dietary polyphenols against metabolic disorders: how far have we progressed in the understanding of the molecular mechanisms of action of these compounds? Crit Rev Food Sci Nutr 57:1769–1786. https://doi.org/10.1080/10408398.2014.980499

5. Milenkovic D, Morand C, Cassidy A, Konic-Ristic A, Tomás-Barberán F, Ordovas JM, Kroon P, De Caterina R, Rodriguez-Mateos A (2017) Interindividual variability in biomarkers of cardiometabolic health after consumption of major plant-food bioactive compounds and the determinants involved. Adv Nutr 8:558–570. https://doi.org/10.3945/an.116.013623

6. Manach C, Milenkovic D, Van de Wiele T, Rodriguez-Mateos A, de Roos B, Garcia-Conesa MT, Landberg R, Gibney ER, Heinonen M, Tomás-Barberán F, Morand C (2017) Addressing the inter-individual variation in response to consumption of plant food bioactives: towards a better understanding of their role in healthy aging and cardiometabolic risk reduction. Mol Nutr Food Res 61:1600557. https://doi.org/10.1002/mnfr.201600557

7. Bayram B, González-Sarrías A, Istas G, Garcia-Aloy M, Morand C, Tuohy K, García-Villalba R, Mena P (2018) Breakthroughs in the health effects of plant food bioactives: a perspective on microbiomics, nutri(epi)genomics, and metabolomics. J Agric Food Chem 66:10686–10692. https://doi.org/10.1021/acs.jafc.8b03385

8. Menezes R, Rodriguez-Mateos A, Kaltsatou A, González-Sarrías A, Greyling A, Giannaki C, Andres-Lacueva C, Milenkovic D, Gibney ER, Dumont J, Schär M, Garcia-Aloy M, Palma-Duran SA, Ruskovska T, Maksimova V, Combet E, Pinto P (2017) Impact of flavonols on cardiometabolic biomarkers: a meta-analysis of randomized controlled human trials to explore the role of inter individual variability. Nutrients 9:e117. https://doi.org/10.3390/nu9020117

9. González-Sarrías A, Combet E, Pinto P, Mena P, Dall'Asta M, Garcia-Aloy M, Rodríguez-Mateos A, Gibney ER, Dumont J, Massaro M, Sánchez-Meca J, Morand C, García-Conesa MT (2017) A systematic review and meta-analysis of the effects of flavanol-containing tea, cocoa and apple products on body composition and blood lipids: exploring the factors responsible for variability in their efficacy. Nutrients 9:746. https://doi.org/10.3390/nu9070746

10. García-Conesa MT, Chambers K, Combet E, Pinto P, Garcia-Aloy M, Andrés-Lacueva C, de Pascual-Teresa S, Mena P, Konic Ristic A, Hollands WJ, Kroon PA, Rodríguez-Mateos A, Istas G, Kontogiorgis CA, Rai DK, Gibney ER, Morand C, Espín JC, González-Sarrías A (2018) Meta-analysis of the effects of foods and derived products containing ellagitannins and anthocyanins on cardiometabolic biomarkers: analysis of factors influencing variability of the individual responses. Int J Mol Sci 19:694. https://doi.org/10.3390/ijms19030694

11. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche P, Lang T (2001) The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 34:663–694. https://doi.org/10.7326/0003-4819-134-8-200104170-0001

12. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG, CONSORT Group (2012) CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 10:28–55. https://doi.org/10.1016/j.ijsu.2011.10.001

13. Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C, CONSORT Group (2006) Reporting randomized, controlled trials of herbal interventions: an elaborated CONSORT statement. Ann Intern Med 144:364–367. https://doi.org/10.7326/0003-4819-144-5-200603070-00013

14. Morand C, Tomás-Barberán FA (2019) Interindividual variability in absorption, distribution, metabolism, and excretion of food phytochemicals should be reported. J Agric Food Chem 67:3843–3844. https://doi.org/10.1021/acs.jafc.9b01175

15. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ (1996) Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 17:1–12

16. Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, Knipschild PG (1998) The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 51:1235–1241

17. van Tulder M, Furlan A, Bombardier C, Bouter L, Editorial Board of the Cochrane Collaboration Back Review Group (2003) Updated method guidelines for systematic reviews in the Cochrane collaboration back review group. Spine 28:1290–1299

18. Berger VW, Alperson SY (2009) A general framework for the evaluation of clinical trial quality. Rev Recent Clin Trials 4:79–88

19. Page MJ, McKenzie JE, Higgins JPT (2018) Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. BMJ Open 8:e019703. https://doi.org/10.1136/bmjopen-2017-019703

20. Gabler NB, Duan N, Liao D, Elmore JG, Ganiats TG, Kravitz RL (2009) Dealing with heterogeneity of treatment effects: is the literature up to the challenge? Trials 10:43. https://doi.org/10.1186/1745-6215-10-43

21. Fernandez Y, Garcia E, Nguyen H, Duan N, Gabler NB, Kravitz RL (2010) Assessing heterogeneity of treatment effects: are authors misinterpreting their results? Health Serv Res 45:283–301. https://doi.org/10.1111/j.1475-6773.2009.01064.x

22. Dahan M, Scemama C, Porcher R, Biau DJ (2018) Reporting of heterogeneity of treatment effect in cohort studies: a review of the literature. BMC Med Res Methodol 18:10. https://doi.org/10.1186/s12874-017-0466-6

23. Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA (2010) Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. Trials 11:85. https://doi.org/10.1186/1745-6215-11-85

24. Pocock SJ, McMurray JJV, Collier TJ (2015) Statistical controversies in reporting of clinical trials: part 2 of a 4-part series on

statistics for clinical trials. J Am Coll Cardiol 66:2648–2662. https://doi.org/10.1016/j.jacc.2015.10.023

25. Assmann SF, Pocock SJ, Enos LE, Kasten LE (2000) Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 355:1064–1069

26. Pocock SJ, Gersh BJ (2014) Do current clinical trials meet society's needs?: a critical review of recent evidence. J Am Coll Cardiol 64:1615–1628. https://doi.org/10.1016/j.jacc.2014.08.008

27. Moher D, Schulz KF, Simera I, Altman DG (2010) Guidance for developers of health research reporting guidelines. PLoS Med 7:e1000217. https://doi.org/10.1371/journal.pmed.1000217

28. Whiting P, Wolff R, Mallett S, Simera I, Savović J (2017) A proposed framework for developing quality assessment tools. Syst Rev 6:204. https://doi.org/10.1186/s13643-017-0604-6

29. World Health Organization (2014) WHO handbook for guideline development, 2nd edn. World Health Organization. http://www.who.int/iris/handle/10665/145714. Accessed 2 May 2016

30. Borel P, Desmarchelier C, Nowicki M, Bott R, Morange S, Lesavre N (2014) Interindividual variability of lutein bioavailability in healthy men: characterization, genetic variants involved, and relation with fasting plasma lutein concentration. Am J Clin Nutr 100:168–175. https://doi.org/10.3945/ajcn.114.085720

31. Setchell KD, Cole SJ (2006) Method of defining equol-producer status and its frequency among vegetarians. J Nutr 136:2188–2193

32. Heiss C, Sansone R, Karimi H, Krabbe M, Schuler D, Rodriguez-Mateos A, Kraemer T, Cortese-Krott MM, Kuhnle GG, Spencer JP, Schroeter H, Merx MW, Kelm M, FLAVIOLA Consortium, European Union 7th Framework Program (2015) Impact of cocoa flavanol intake on age-dependent vascular stiffness in healthy men: a randomized controlled, double-masked trial. Age (Dordr) 37:9794. https://doi.org/10.1007/s11357-015-9794-9

33. Sanchez-Muniz FJ, Maki KC, Schaefer EJ, Ordovas JM (2009) Serum lipid and antioxidant responses in hypercholesterolemic men and women receiving plant sterol esters vary by apolipoprotein E genotype. J Nutr 139:13–19. https://doi.org/10.3945/jn.108.090969

34. Borel P, Desmarchelier C, Nowicki M, Bott R (2015) Lycopene bioavailability is associated with a combination of genetic variants. Free Radic Biol Med 83:238–244. https://doi.org/10.1016/j.freeradbiomed.2015.02.033

35. Song KB, Atkinson C, Frankenfeld CL, Jokela T, Wähälä K, Thomas WK, Lampe JW (2006) Prevalence of daidzein-metabolizing phenotypes differs between Caucasian and Korean American women and girls. J Nutr 136:1347–1351

36. Ibero-Baraibar I, Abete I, Navas-Carretero S, Massis-Zaid A, Martinez JA, Zulet MA (2014) Oxidised LDL levels decreases after the consumption of ready-to-eat meals supplemented with cocoa extract within a hypocaloric diet. Nutr Metab Cardiovasc Dis 24:416–422. https://doi.org/10.1016/j.numecd.2013.09.017

37. Strandhagen E, Zetterberg H, Aires N, Palmér M, Rymo L, Blennow K, Thelle DS (2004) The apolipoprotein E polymorphism and the cholesterol-raising effect of coffee. Lipids Health Dis 3:26

38. Gasper AV, Al-Janobi A, Smith JA, Bacon JR, Fortun P, Atherton C, Taylor MA, Hawkey CJ, Barrett DA, Mithen RF (2005) Glutathione S-transferase M1 polymorphism and metabolism of sulforaphane from standard and high-glucosinolate broccoli. Am J Clin Nutr 82:1283–1291

39. Steck SE, Gammon MD, Hebert JR, Wall DE, Zeisel SH (2007) GSTM1, GSTT1, GSTP1, and GSTA1 polymorphisms and urinary isothiocyanate metabolites following broccoli consumption in humans. J Nutr 137(4):904–909

40. Kreijkamp-Kaspers S, Kok L, Bots ML, Grobbee DE, van der Schouw YT (2004) Dietary phytoestrogens and vascular function in postmenopausal women: a cross-sectional study. J Hypertens 22:1381–1388

41. Strandhagen E, Zetterberg H, Aires N, Palmér M, Rymo L, Blennow K, Landaas S, Thelle DS (2004) The methylenetetrahydrofolate reductase C677T polymorphism is a major determinant of coffee-induced increase of plasma homocysteine: a randomized placebo controlled study. Int J Mol Med 13(6):811–815

42. Selma MV, Romo-Vaquero M, García-Villalba R, González-Sarrías A, Tomás-Barberán FA, Espín JC (2016) The human gut microbial ecology associated with overweight and obesity determines ellagic acid metabolism. Food Funct 7:1769–1774. https://doi.org/10.1039/c5fo01100k

43. Fisher ND, Hollenberg NK (2006) Aging and vascular responses to flavanol-rich cocoa. J Hypertens 24:1575–1580

44. Renda G, Zimarino M, Antonucci I, Tatasciore A, Ruggieri B, Bucciarelli T, Prontera T, Stuppia L, De Caterina R (2012) Genetic determinants of blood pressure responses to caffeine drinking. Am J Clin Nutr 95:241–248. https://doi.org/10.3945/ajcn.111.018267

45. Rideout TC, Harding SV, Mackay D, Abumweis SS, Jones PJ (2010) High basal fractional cholesterol synthesis is associated with nonresponse of plasma LDL cholesterol to plant sterol therapy. Am J Clin Nutr 92:41–46. https://doi.org/10.3945/ajcn.2009.29073

46. Rodriguez-Mateos A, Cifuentes-Gomez T, Gonzalez-Salvador I, Ottaviani JI, Schroeter H, Kelm M, Heiss C, Spencer JP (2015) Influence of age on the absorption, metabolism, and excretion of cocoa flavanols in healthy subjects. Mol Nutr Food Res 59:1504–1512. https://doi.org/10.1002/mnfr.201500091

47. Egert S, Bosy-Westphal A, Seiberl J, Kürbitz C, Settler U, Plachta-Danielzik S, Wagner AE, Frank J, Schrezenmeir J, Rimbach G, Wolffram S, Müller MJ (2009) Quercetin reduces systolic blood pressure and plasma oxidised low-density lipoprotein concentrations in overweight subjects with a high-cardiovascular disease risk phenotype: a double-blinded, placebo-controlled cross-over study. Br J Nutr 102:1065–1074. https://doi.org/10.1017/S0007114509359127

48. Ostertag LM, Kroon PA, Wood S, Horgan GW, Cienfuegos-Jovellanos E, Saha S, Duthie GG, de Roos B (2013) Flavan-3-ol-enriched dark chocolate and white chocolate improve acute measures of platelet function in a gender-specific way-a randomized-controlled human intervention trial. Mol Nutr Food Res 57:191–202. https://doi.org/10.1002/mnfr.201200283

49. Zhao HL, Houweling AH, Vanstone CA, Jew S, Trautwein EA, Duchateau GS, Jones PJ (2008) Genetic variation in ABC G5/G8 and NPC1L1 impact cholesterol response to plant sterols in hypercholesterolemic men. Lipids 43:1155–1164. https://doi.org/10.1007/s11745-008-3241-y

50. González-Barrio R, Borges G, Mullen W, Crozier A (2010) Bioavailability of anthocyanins and ellagitannins following consumption of raspberries by healthy humans and subjects with an ileostomy. J Agric Food Chem 58:3933–3939. https://doi.org/10.1021/jf100315d

51. Tomás-Navarro M, Vallejo F, Sentandreu E, Navarro JL, Tomás-Barberán FA (2014) Volunteer stratification is more relevant than technological treatment in orange juice flavanone bioavailability. J Agric Food Chem 62:24–27. https://doi.org/10.1021/jf4048989

52. Mackay DS, Gebauer SK, Eck PK, Baer DJ, Jones PJ (2015) Lathosterol-to-cholesterol ratio in serum predicts cholesterol-lowering response to plant sterol consumption in a dual-center, randomized, single-blind placebo-controlled trial. Am J Clin Nutr 101:432–439. https://doi.org/10.3945/ajcn.114.095356

53. Usui T, Tochiya M, Sasaki Y, Muranaka K, Yamakage H, Himeno A, Shimatsu A, Inaguma A, Ueno T, Uchiyama S, Satoh-Asahara N (2013) Effects of natural S-equol supplements on overweight or obesity and metabolic syndrome in the Japanese, based on sex and equol status. Clin Endocrinol (Oxf) 78:365–372. https://doi.org/10.1111/j.1365-2265.2012.04400.x

54. Bolca S, Possemiers S, Maervoet V, Huybrechts I, Heyerick A, Vervarcke S, Depypere H, De Keukeleire D, Bracke M, De Henauw S, Verstraete W, Van de Wiele T (2007) Microbial and dietary factors associated with the 8-prenylnaringenin producer phenotype: a dietary intervention trial with fifty healthy post-menopausal Caucasian women. Br J Nutr 98:950–959

55. Setchell KD, Brown NM, Summer S, King EC, Heubi JE, Cole S, Guy T, Hokin B (2013) Dietary factors influence production of the soy isoflavone metabolite s-(-)equol in healthy adults. J Nutr 143:1950–1958. https://doi.org/10.3945/jn.113.179564

56. Kuijsten A, Arts IC, Vree TB, Hollman PC (2005) Pharmacokinetics of enterolignans in healthy men and women consuming a single dose of secoisolariciresinol diglucoside. J Nutr 135:795–801

57. Wang TT, Edwards AJ, Clevidence BA (2013) Strong and weak plasma response to dietary carotenoids identified by cluster analysis and linked to beta-carotene 15,15′-monooxygenase 1 single nucleotide polymorphisms. J Nutr Biochem 24:1538–1546. https://doi.org/10.1016/j.jnutbio.2013.01.001

58. Miller RJ, Jackson KG, Dadd T, Mayes AE, Brown AL, Minihane AM (2011) The impact of the catechol-O-methyltransferase genotype on the acute responsiveness of vascular reactivity to a green tea extract. Br J Nutr 105:1138–1144. https://doi.org/10.1017/S0007114510004836

59. West SG, McIntyre MD, Piotrowski MJ, Poupin N, Miller DL, Preston AG, Wagner P, Groves LF, Skulas-Ray AC (2014) Effects of dark chocolate and cocoa consumption on endothelial function and arterial stiffness in overweight adults. Br J Nutr 111:653–661. https://doi.org/10.1017/S0007114513002912

60. Mark DB, Lee KL, Harrell FE Jr (2016) Understanding the role of p values and hypothesis tests in clinical research. JAMA Cardiol 1:1048–1054. https://doi.org/10.1001/jamacardio.2016.3312

61. Doshi P, Goodman SN, Ioannidis JP (2013) Raw data from clinical trials: within reach? Trends Pharmacol Sci 34:645–647. https://doi.org/10.1016/j.tips.2013.10.006

62. European Commission (2016) H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. Accessed 25 July 2018

63. Ohmann C, Banzi R, Canham S et al (2017) Sharing and reuse of individual participant data from clinical trials: principles and recommendations. BMJ Open 7:e018647. https://doi.org/10.1136/bmjopen-2017-018647

64. Field AP, Miles J, Field Z (2012) Discovering statistics using R. SAGE, London

65. Ghasemi A, Zahediasl S (2012) Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab 10:486–489. https://doi.org/10.5812/ijem.3505

66. Thode H (2002) Testing for Normality. CRC Press, Boca Raton. https://doi.org/10.1201/9780203910894

67. Sullivan GM, Feinn R (2012) Using effect size-or why the p value is not enough. J Grad Med Educ 4:279–282. https://doi.org/10.4300/JGME-D-12-00156.1

68. Durlak JA (2009) How to select, calculate, and interpret effect sizes. J Pediatr Psychol 3:917–928. https://doi.org/10.1093/jpepsy/jsp004

69. Weissgerber TL, Milic NM, Winham SJ, Garovic VD (2015) Beyond bar and line graphs: time for a new data presentation paradigm.

PLoS Biol 13:e1002128. https://doi.org/10.1371/journal.pbio.1002128

70. Brindani N, Mena P, Calani L, Benzie I, Choi SW, Brighenti F, Zanardi F, Curti C, Del Rio D (2017) Synthetic and analytical strategies for the quantification of phenyl-γ-valerolactone conjugated metabolites in human urine. Mol Nutr Food Res 61:e1700077. https://doi.org/10.1002/mnfr.201700077

71. Jüni P, Altman DG, Egger M (2001) Systematic reviews in health care: assessing the quality of controlled clinical trials. BMJ 323:42–46. https://doi.org/10.1136/bmj.323.7303.42

72. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF (1996) Improving the quality of reporting of randomized controlled trials: the CONSORT statement. JAMA 276:637–639

73. Schulz KF, Altman DG, Moher D; CONSORT Group (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 52:726–732. https://doi.org/10.7326/0003-4819-152-11-201006010-00232

74. Kasenda B, Schandelmaier S, Sun X, von Elm E, You J, Blümle A, Tomonaga Y, Saccilotto R, Amstutz A, Bengough T, Meerpohl JJ, Stegert M, Olu KK, Tikkinen KA, Neumann I, Carrasco-Labra A, Faulhaber M, Mulla SM, Mertz D, Akl EA, Bassler D, Busse JW, Ferreira-González I, Lamontagne F, Nordmann A, Gloy V, Raatz H, Moja L, Rosenthal R, Ebrahim S, Vandvik PO, Johnston BC, Walter MA, Burnand B, Schwenkglenks M, Hemkens LG, Bucher HC, Guyatt GH, Briel M, DISCO Study Group (2014) Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications. BMJ 349:g4539. https://doi.org/10.1136/bmj.g4539

75. International Committee of Medical Journal Editors (2018) Recommendations for the conduct, reporting, editing and publication of scholarly work in medical journals. http://www.ICMJE.org

76. Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G (2014) How to use a subgroup analysis: users' guide to the medical literature. JAMA 311:405–411. https://doi.org/10.1001/jama.2013.285063

77. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM (2007) Statistics in medicine—reporting of subgroup analyses in clinical trials. N Engl J Med 357:2189–2194

78. Garcia-Aloy M, Andres-Lacueva C (2018) Food intake biomarkers for increasing the efficiency of dietary pattern assessment through the use of metabolomics: unforeseen research requirements for addressing current gaps. J Agric Food Chem 66(1):5–7. https://doi.org/10.1021/acs.jafc.7b05586

79. 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://doi.org/10.1002/mnfr.201701064

80. Collins F, Varmus H (2015) A new initiative on precision medicine. N Engl J Med 372(9):793–795

81. Khoury M, Iademarco M, Riley W (2016) Precision public health for the era of precision medicine. Am J Prev Med 50(3):398–401. https://doi.org/10.1016/j.amepre.2015.08.031

82. National Academies of Sciences, Engineering, and Medicine (2016) Interindividual variability: new ways to study and implications for decision making: workshop in bries. The National Academies Press, Washington, DC. https://doi.org/10.17226/23413

Affiliations

Marina Nikolic[1] · Aleksandra Konic Ristic[1,2] · Antonio González-Sarrías[3] · Geoffrey Istas[4] ·
Mireia Urpi-Sarda[5,6] · Margherita Dall'Asta[7] · Laurent-Emmanuel Monfoulet[8] · Lieselotte Cloetens[9] ·
Banu Bayram[10] · Maria Rosaria Tumolo[11] · Mihail Chervenkov[12,13] · Egeria Scoditti[14] · Marika Massaro[14] ·
Noemi Tejera[15] · Desislava Abadjieva[16] · Karen Chambers[17] · Irena Krga[1] · Francisco A. Tomás-Barberán[3] ·
Christine Morand[8] · Rodrigo Feliciano[18] · Rocío García-Villalba[3] · Mar Garcia-Aloy[5,6] · Pedro Mena[7]

1. Institute for Medical Research, University of Belgrade, Belgrade, Serbia

2. UCD Institute of Food and Health, University College Dublin, Belfield, Dublin, Ireland

3. Laboratory of Food and Health, Research Group on Quality, Safety and Bioactivity of Plant Foods, CEBAS-CSIC, Murcia, Spain

4. Department of Nutritional Sciences, Faculty of Life Sciences and Medicine, School of Life Course Sciences, King's College London, London, UK

5. Biomarkers and Nutrimetabolomic Laboratory, Department of Nutrition, Food Sciences and Gastronomy, XaRTA, INSA, Faculty of Pharmacy and Food Sciences, University of Barcelona, Santa Coloma De Gramenet, Spain

6. CIBER de Fragilidad y Envejecimiento Saludable (CIBERFES), Instituto de Salud Carlos III, Barcelona, Spain

7. Human Nutrition Unit, Department of Food and Drugs, University of Parma, Medical School Building C, Via Volturno, 39, 43125 Parma, Italy

8. Unité de Nutrition Humaine (UNH), Institut National de la Recherche Agronomique (INRA), Université Clermont Auvergne, CRNH Auvergne, Clermont-Ferrand, France

9. Biomedical Nutrition, Pure and Applied Biochemistry, Lund University, Lund, Sweden

10. Department of Nutrition and Dietetics, University of Health Sciences, Istanbul, Turkey

11. Research Unit of Brindisi, Institute for Research on Population and Social Policies, National Research Council, Brindisi, Italy

12. Faculty of Veterinary Medicine, University of Forestry, Sofia, Bulgaria

13. Institute of Neurobiology, Bulgarian Academy of Sciences, Sofia, Bulgaria

14. Institute of Clinical Physiology (IFC), National Research Council (CNR), Lecce, Italy

15. Department of Nutrition and Preventive Medicine, Norwich Medical School, University of East Anglia, Norwich, UK

16. Institute of Biology and Immunology of Reproduction, Bulgarian Academy of Sciences, Sofia, Bulgaria

17. Quadram Institute Bioscience, Norwich Research Park, Norwich, UK

18. Division of Cardiology, Pulmonology, and Vascular Medicine, Medical Faculty, University of Duesseldorf, Dusseldorf, Germany