Evaluating the Applicability of Data-Driven Dietary Patterns to Independent Samples with a Focus on Measurement Tools for Pattern Similarity.

Adela Castelló, Brian Buijsse, Miguel Martín, Amparo Ruiz, Ana M. Casas, Jose M. Baena-Cañada, Roberto Pastor-Barriuso, Silvia Antolín, Manuel Ramos, Monserrat Muñoz, Ana Lluch, Ana de Juan-Ferré, Carlos Jara, Virginia Lope, María A. Jimeno, Esperanza Arriola-Arellano, Elena Díaz, Vicente Guillem, Eva Carrasco, Beatriz Pérez-Gómez, Jesús Vioque, Marina Pollán.

J Acad Nutr Diet. 2016 Dec;116(12):1914-1924.e6.

which has been published in final form at https://doi.org/10.1016/j.jand.2016.05.008

Title: Evaluating the applicability of data-driven dietary patterns to independent samples with focus on measurement tools for pattern similarity.

Keywords: Dietary patterns; Pattern Similarity; Congruence Coefficient; Breast Cancer; Mediterranean/Western Patterns.

Word Count: 3673

Authors: Adela Castelló[1,2,3] PhD; Brian Buijsse[3] PhD; Miguel Martín[4,5] PhD; Amparo Ruiz[6] PhD; Ana M. Casas[7] PhD; Jose M. Baena-Cañada[8] PhD; Roberto Pastor[1,2] PhD; Silvia Antolín[9] MD; Manuel Ramos[10] MD; Monserrat Muñoz[11] PhD; Ana Lluch[12] PhD; Ana de Juan-Ferré[13] PhD; Carlos Jara[14]PhD; Virginia Lope[1,2] PhD; María A. Jimeno[15] BSc; Esperanza Arriola-Arellano[8]MD; Elena Díaz[7] PhD; Vicente Guillem[6] PhD; Eva Carrasco[15] PhD; Beatriz Pérez-Gómez[1,2] PhD; Jesús Vioque[2,16] PhD; Marina Pollán[1,2] PhD on behalf of GEICAM researchers^.

^All credentials were issued in Spain

[1] Cancer Epidemiology Unit, National Center for Epidemiology, Instituto de Salud Carlos III, Av Monforte de Lemos, 5, 28029 Madrid, Spain.

[2] Consortium for Biomedical Research in Epidemiology & Public Health (CIBERESP), Instituto de Salud Carlos III, Madrid, Spain.

[3] Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbrücke, Arthur-Scheunert-Allee 114-116, 14558 Nuthetal, Germany

[4] Medical Oncology Unit. Hospital Clínico Universitario San Carlos. Calle Profesor Martín Lagos, 28040 Madrid, Spain.

[5] Health Research Institute Gregorio Marañón, Universidad Complutense, Calle Doctor Esquerdo, 46, 28007 Madrid, Spain.

[6] Medical Oncology Unit. Instituto Valenciano de Oncología. Calle del Profesor Beltrán Bàguena, 8, 46009 Valencia, Spain.

[7] Medical Oncology Unit. Hospital Virgen del Rocío. Av Manuel Siurot, s/n, 41013 Sevilla, Spain.

[8] Medical Oncology Unit. Hospital Puerta del Mar. Av Ana de Viya, 21, 11009 Cádiz, Spain.

[9] Medical Oncology Unit. Complejo Hospitalario Universitario. Calle xubias de arriba, 84, 15006 A Coruña, Spain.

[10] Medical Oncology Unit. Centro Oncológico de Galicia. Av Montserrat,s/n, 15009 A Coruña, Spain.

[11] Medical Oncology Unit. Hospital Clinic I Provincial. Carrer Villarroel, 170, 08036 Barcelona, Spain.

[12] Hematology-Oncology Department. Hospital Clínico Universitario de Valencia / INCLIVA, Universidad de Valencia, Avinguda de Blasco Ibáñez, 17, 46010 Valencia, Spain.

[13] Medical Oncology Unit. Hospital Marqués de Valdecilla. Av Valdecilla, s/n, 39008 Santander, Spain.

[14] Medical Oncology Unit. Fundación Hospitalaria de Alcorcón. Calle Budapest, 1, 28922 Alcorcón, Spain.

[15] GEICAM (Spanish Breast Cancer Group), Avda. de los Pirineos nº 7, 28703 San Sebastián de los Reyes, Madrid, Spain.

[16] Department of Public Health. Universidad Miguel Hernandez. Carretera N - 332, s/n, 03550 Sant Joan D'Alacant, Spain.

Email Addresses:

Adela Castelló: acastello@isciii.es

Brian Buijsse: brian.buijsse@gmail.com

Miguel Martín: mmartin@geicam.org

Amparo Ruiz: aruiz@fico.org

Ana M. Casas: anacasas05@gmail.com

Jose M Baena-Cañada: jmbaenacanada@gmail.com

Roberto Pastor: rpastor@isciii.es

Silvia Antolín: silviaantolin@hotmail.com

Manuel Ramos: manuel.ramos@cog.es

Montserrat Muñoz: mmunoz@clinic.ub.es

Ana Lluch: lluch_ana@gva.es

Ana de Juan-Ferré: ajuan@humv.es

Carlos Jara: cjara@fhalcorcon.es

Virginia Lope: vicarvajal@isciii.es

María A. Jimeno: majimeno@geicam.org

Esperanza Arriola-Arellano: arriolaarellano@hotmail.com

Elena Díaz: elediazp@gmail.com

Vicente Guillem: vguillem@fivo.org

Eva Carrasco: evacarrasco@geicam.org

Beatriz Pérez-Gómez: bperez@isciii.es

Jesús Vioque: vioque@umh.es

Marina Pollán: mpollan@isciii.es

CORRESPONDING AUTHOR:

Dr. Adela Castelló

Cancer Epidemiology Unit, National Center for Epidemiology.

Instituto de Salud Carlos III.

Avenida Monforte de Lemos, 5, 28029, Madrid, Spain.

Phone: +34 91 822 2667.

Fax: +34 91 387 7815.

e-mail: acastello@isciii.es

ABSTRACT

Background: Diet is a key modifiable risk for many chronic diseases. But it remains unclear if dietary patterns from one study sample are generalizable to other independent populations.

Objective: The primary objective of this study was to assess whether data-driven dietary patterns from one study sample are applicable to other populations. The secondary objective was to assess the validity of two criteria of pattern similarity.

Methods: Six dietary patterns – "Western" (n=3), "Mediterranean", "Prudent", and "Healthy" – from three published studies on breast cancer were reconstructed in a case-control study of 973 breast cancer cases and 973 controls. Three more "internal" patterns ("Western", "Prudent", "Mediterranean") were derived from this case-control study's own data.

Statistical Analysis: Applicability was assessed by comparing the six reconstructed patterns with the three internal dietary patterns, using the congruence coefficient (CC) between pattern loadings. If any pair met either of two commonly used criteria for declaring patterns similar ($CC \geq 0.85$ or a statistically significant ($p<0.05$) Pearson correlation), then the true similarity of those two dietary patterns was double-checked by comparing their associations to risk for breast cancer, in order to assess whether those two criteria of similarity are actually reliable.

Results: Five of the six reconstructed dietary patterns showed high congruence ($CC>0.9$) to their corresponding dietary pattern derived from the case-control study's data. Similar associations with risk for breast cancer were found in all pairs of dietary patterns that had high CC but not in all pairs of dietary patterns with statistically significant correlations.

Conclusions: Similar dietary patterns can be found in independent samples. The p-value of a correlation coefficient is less reliable than the CC as a criterion for declaring two dietary patterns

similar. This study shows that diet scores based on a particular study are generalizable to other populations.

INTRODUCTION

Diet is a key modifiable risk factor for many chronic diseases [1-3]. For many years, nutritional epidemiology has focused on individual dietary factors in relation to disease. However, dietary pattern analysis has emerged as an important area of research. The study of dietary patterns may better capture dietary variability in the population than tracking individual foods or nutrients, while also accounting for interactions between dietary factors [4-6]. Many investigator-driven indexes evaluate dietary quality against predefined criteria [7, 8]. Reproducibility and consistency of the associations between the adherence to such indexes and disease have been widely explored [1-3, 8, 9]. Nevertheless, investigator-driven dietary patterns are applicable only in populations that consume the foods described in the index and its construction is mainly based on the existing evidence of the association between diet and cardiovascular disease, making them less than ideal to explore associations with other diseases [4-6].

Dietary patterns that are more representative of a specific population can be identified with data-driven methods like principal component analysis (PCA), factor analysis (FA) and cluster analysis ("a posteriori" methods) [10]. Data-driven dietary patterns also present the advantage of being extracted independently of disease associations, which allows evaluation of the role of actual eating habits in disease risk. However, one of the main criticisms of these methods is that the patterns extracted are dependent on the population and, therefore, difficult to apply to other settings [6, 11, 12]. Conversely, some authors have proposed methods to construct simplified measures of dietary patterns that may facilitate their replication in different populations [13]. To our knowledge, no studies have explored the applicability of data-driven dietary patterns using the simplified measures to date. Despite the fact that various authors have proposed methods to evaluate the congruence between components or factors extracted with PCA or FA [14-16], such congruence is usually assessed with simple linear correlations between adherence scores, basing the conclusion about pattern similarity only on the significance of such correlations [17-20].

The objective of this study was to assess whether data-driven dietary patterns extracted in different populations are applicable to a sample of participants of similar characteristics, comparing different measurements of similarity of patterns and their associations with BC risk. This was achieved by re-constructing dietary patterns from other populations and comparing their characteristics and associations with breast cancer against similarly labeled dietary patterns that were internally derived with PCA in a case-control study of breast cancer.

MATERIALS AND METHODS

EpiGEICAM study population

Data used were from the EpiGEICAM study, whose design description has been provided previously [21]. Briefly, EpiGEICAM is a Spanish case-control study that recruited, between 2006 and 2011, *1017* incident cases of female breast cancer (BC) diagnosed in the Oncology departments of *23* hospitals affiliated with the Spanish Breast Cancer Group (GEICAM). Each case was matched with a healthy control of similar age (± 5 years), selected from cases' in-laws, friends, neighbors or work colleagues residing in the same town. Cases and controls completed a structured questionnaire on demographic and anthropometric characteristics, personal, family, gynecological, obstetric and occupational history, past physical activity and diet. Dietary intake in the last five years was estimated using a *117*-item semi-quantitative food frequency questionnaire (FFQ) [22] adapted to and validated in different Spanish adult populations [23, 24]. Postmenopausal status was defined as absence of menstruation in the last 12 months.

The EpiGEICAM study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects were approved by the Ethics Committees of the *23* participating hospitals. Written informed consent was obtained from all subjects.

Dietary patterns in EpiGEICAM

Three dietary patterns that characterize the diet of the Spanish women have been recently identified in the control group of EpiGEICAM study [21] using PCA: The first pattern was labeled *Western* and characterized by high intake of high-fat dairy products, processed meat, refined grains, sweets, caloric drinks, convenience food and sauces, and by low intake of low-fat dairy products and whole grains; high adherence to this pattern was associated with an increased risk of BC. The second was labeled *Prudent*, characterized by high intake of low-fat dairy products, fruits, vegetables, whole grains and juices; this pattern was not associated with BC. The third pattern was labeled *Mediterranean* because it was characterized by high intake of fish, vegetables, legumes, boiled potatoes, fruits, olives and vegetable oil, and by low intake of juices. A strong adherence to this pattern was associated with lower BC risk.

Dietary patterns in independent populations

To assess the applicability of data-driven dietary patterns developed in different populations with similar characteristics, a bibliographic search of the scientific literature published between 2000 and 2014 and reporting on the association between dietary patterns and BC risk was carried out. The search was performed in PubMed using the following keywords: Breast Neoplasms (Mesh term), diet patterns, dietary patterns, and food patterns. Additionally, all references included in three recent reviews [25-27] were screened. Eligibility criteria were the following:

1) The study population consisted of Caucasian adult women;

2) Dietary patterns were derived with PCA or FA;

3) The study reported pattern loadings ≥|0.15| for food groups;

4) Dietary intake was classified in food groups that allowed the replication of dietary patterns in EpiGEICAM data.

Of the 44 identified articles, 3 were eligible for inclusion. Six dietary patterns from these studies were selected: the *Western* and *Mediterranean* dietary patterns from Bessaoud et al. (France) [28], the *Western* and *Prudent* patterns from Adebamowo et al. (USA) [29] and the *Western* and *Healthy* patterns in Terry et al.(Sweden) [30]. The following patterns were compared;

1. Castelló's Western with Bessaoud's, Adebamowo's and Terry`s Western;
2. Castello's Prudent and Mediterranean with Bessaoud's Mediterranean, Adebamowo's Prudent and Terry`s Healthy.
3. Castello's Mediterranean with Bessaoud's Mediterranean, Adebamowo's Prudent and Terry`s Healthy.

Given that the differences between dietary habits identified under the names of *Mediterranean/Prudent/Healthy* are often subtle, both, Castelló's *Prudent* and *Mediterranean*, were compared with Bessaoud's *Mediterranean*, Adebamowo's *Prudent* and Terry`s *Healthy*. A description of these studies is provided in the supplementary material (**Supplementary Table 1**).

Applicability

The PCA reports, for a given pattern, a set of weights associated to each food group (commonly called component/pattern weights) that is used to calculate pattern scores, defined, for each individual, as a weighted sum of the food group consumption. Pattern scores measure the extent of compliance with the pattern[31]. Afterwards, these scores are correlated with the food group consumption to calculate the pattern loadings, which indicate the importance of individual food groups in each pattern. It is important to note that pattern weights and pattern loadings give similar information, except that they are measured on different scales (weights are standardized into Z score form). Since usually only pattern loadings are given in articles constructing data-driven dietary patterns with PCA or FA, the pattern loadings will be used to compute pattern scores in order to assess similarity between patterns:

1. Food consumption (in grams) collected within EpiGEICAM study, was grouped into the food groups defined by Bessaoud et al. [28], Adebamowo et al. [29] and Terry et al. [30] in their original articles. Items included in each of the patterns are summarized in Supplementary Table 2.

2. Since only information of pattern loadings is usually provided, and taking into account that weights and loadings give similar information, pattern scores of adherence were calculated as the linear combination of the consumption of the food groups constructed in step 1, weighted by the original pattern loadings reported by these studies (**Table 1**). Given that most studies present the component loadings only when those are over a certain threshold (often ≥|0.15|) only food groups whose component loadings were ≥|0.15|, were considered:

$$P_{ki} = \sum_{j:|L_{kj}|\geq 0.15} L_{kj} \cdot C_{ji}$$

P = Pattern Score; L= Pattern Loading; C= Centered food group consumption
k= 1,...,9 for Castelló et al. Western, Prudent and Mediterranean; Bessaoud et al. Western and Mediterranean; Adebamowo et al. Western and Prudent; and Terry et al. Western and Healthy;
i = 1,...,1946 women
j=1,..., s food groups (s<26)

As a first measure to assess the similarity of pairs of patterns, the Pearson's correlation coefficients (Corr) were calculated between the scores of those patterns considered comparable. Traditionally, all correlations that achieve statistical significance are considered an indicator of pattern similarity [17-20, 28, 32-35].

3. The second measure of similarity is the Congruence Coefficient (CC), which is computed using the pattern loadings. However, direct comparison of the original loadings between studies was not possible given the differences in food grouping among them (**Table 1**). In order to obtain pattern loadings associated to the same exact food groups, loadings were recalculated using the food definition provided by Castello et al.[21]. In agreement with their methodological definition [36], pattern loadings were recalculated by correlating the food group consumption of the 26 groups defined in Castello et al. [21] with the 9 pattern scores (Castelló's [21] *Western, Prudent* and *Mediterranean*; Bessaoud's [28] *Western* and *Mediterranean*; Adebamowo's [29] *Western* and *Prudent*; Terry's [30] *Western* and *Healthy*) calculated with the food groups and loadings reported in the

original studies as explained in step 2 (**Tables 1** and **2**). The reconstructed pattern loadings for standard groups were represented graphically (**Figure 1** and **Figure2**). Following the same methodology used in Castelló et al. [21], food groups with a correlation ≥|0.30| were considered to meaningfully contribute to a certain pattern.

After obtaining comparable loadings, the congruence coefficients (CC) between pairs of patterns were calculated. The CCs between pattern 1 (Castelló et .al) and pattern 2 (Bessaoud's, Adebamowo's and Terry's) were calculated [31, 37] as follows:

$$CC = \frac{\sum_{j=1}^{26} l_{1j} \cdot l_{2j}}{\sqrt[2]{\left(\sum_{j=1}^{26} l_{1j}^2\right) \cdot \left(\sum_{j=1}^{26} l_{2j}^2\right)}}$$

l_{1j} and l_{2j} the corresponding loadings for each pattern=1,2 and j=1…26 the different food groups.

CC represents the correlation between pattern loadings based on their deviations from 0 and it is the preferred measure for component/factor similarity extracted with PCA/FA[14]. CC ranges from -1 to 1, a value in the range [0.85-0.94] corresponds to a fair similarity, while a value higher or equal to 0.95 implies that the two compared components/factors can be considered equivalent [14, 15].

An example of the calculations carried out in steps 1-3 is given in the supplementary material using Castelló et al. and Bessaoud et al. definitions of Western pattern (**Supplementary Example 1**).

4. Finally, the associations between patterns and BC risk were calculated by means of separate conditional logistic regression models, one for each of the 9 simplified scores. The scores were included in these models as categorical variables (quartiles of adherence) and also as a continuous term (1sd increase). All models were adjusted by total energy intake; alcohol consumption; body mass index (BMI) from self-reported weight and height (BMI=Kg/m^2); physical activity in the last year; smoking; education; history of breast disease other than cancer; family history of BC; age at menarche; age at first delivery; and menopausal status. The magnitude, direction and significance

of the associations found (**Table 2**) were compared between patterns and against the determination of pattern similarity to explore both, pattern similarity and the adequacy of the Corr and CC to evaluate pattern similarity.

Missing data

BMI (10%), physical activity in the last year (8 %), age at first delivery (5%), smoking habit (<1%), education (<1%) and age at menarche (<1%) contained missing values. As explained in Castelló et al. [21], missing values for these variables were imputed using multiple imputation with chained equations, creating five imputed data sets that were used for subsequent analyses. The final effect is a weighted average of the effects found in these five datasets [38-40].

Analyses were performed using STATA/MP (version 14.0, 2015, StataCorp LP).

RESULTS

After excluding *44* case-control pairs (*n=88*) with incomplete data on diet or implausible reported energy intakes (<750 or >4500 kcal/day) in either the case or the control, final analyses were based on *973* cases-control pairs. Characteristics of the population and dietary patterns identified have been previously described [21].

Comparison of Western patterns composition:

Figure 1 shows the correlation of each food group with the simplified version of the *Western* pattern scores calculated using the loadings published in the four explored studies: Castelló et al., Bessaoud et al., Adebamowo el al. and Terry et al. All of them presented high correlations with the following groups: high-fat dairy, red and processed meat, refined grains, sweets, caloric drinks and convenience food and sauces. However, food grouping from the Bessaoud et al. study showed some important differences: These authors did not take into account other high-fat dairy products than cheese and did not create a category of caloric drinks (two very important components of the

Western pattern) and, in the cereals category, they mixed refined with whole grain (**Table 1**). In spite of this, the correlation between the group of refined grains and Bessaoud's *Western* score was high (**Figure 1)**. However, the correlations with the dairy products and caloric drinks groups as well as the congruence with Castello's *Western* pattern were diminished ($r_{\text{hight-fat dairy}}=0.35$; $r_{\text{caloric drinks}}=0.32$; CC=0.82;) in comparison with Adebamowo's ($r_{\text{hight-fat dairy}}=0.44$; $r_{\text{caloric drinks}}=0.53$; CC=0.92) and Terry's. *Western* scores ($r_{\text{hight-fat dairy}}=0.55$; $r_{\text{caloric drinks}}=0.64$; CC=0.94), which showed a high congruence.

Comparison of Prudent, Healthy and Mediterranean patterns composition:

Similar comparisons between original (**Table 1**) and reproduced scores (**Figure 2**) can be made for *Prudent/Mediterranean/Healthy* patterns. Castelló's Prudent and Mediterranean patterns (shown in the two first columns) shared a high consumption of some items such as fruit and vegetables. However, women following a *Prudent* pattern tend to consume low-fat products, such as low-fat dairy or fruit juices, while women with a high compliance with the Mediterranean pattern eat a greater amount of all types of fish (especially oily fish), legumes, nuts and olive oil. While all three of Bessaoud's *Mediterranean*, Adebamowo's *Prudent* and Terry's *Healthy* loaded high in foods characteristic of the *Mediterranean* diet - such as fish, fruits and vegetables- only Bessaoud's loaded high in olive oil in the original Mediterranean score (Adebamowo et al. and Terry et al. did not create a category for this item, **Table 1**). Subsequently, olive oil showed the greatest correlation in the reproduced version of their pattern (**Figure 2**). On the other hand, Terry's *Healthy* did not have a category for legumes (**Table 1)** and both Adebamowo's *Prudent* and Terry's *Healthy* showed a high correlation with products more typically consumed by women worried about their weight (Castelló's *Prudent*), such as low-fat products or fruit juices in both the original (**Table 1**) and reproduced (**Figure 2**) scores. This was reflected in a higher congruence indicating an identical correspondence of Bessaoud's Mediterranean pattern with Castelló's Mediterranean ($CC_{med}=0.95$); and of Adebamowo's Prudent ($CC_{prud}=0.95$) and Terry's Healthy ($CC_{prud}=0.95$) with Castelló's

Prudent. The congruence with the alternative pattern was weaker for Bessaoud's Mediterranean (CC_{prud}=0.86), Adebamowo's Prudent (CC_{med}=0.88) and Terry's Healthy (CC_{med}=0.77), even if it can be considered fairly high for the first two cases (**Figure 2)**.

Comparison of the associations between the 9 dietary patterns and BC risk:

As expected, all these similarities and dissimilarities between patterns were in consonance with the differences found in their association with BC risk (**Table 2**). The increased risk for the *Western* pattern found with Castello's *Western* (OR_{Q4vsQ1}(95%CI)=1.50(1.09; 2.07)) was not observed for Bessaoud's *Western* (OR_{Q4vsQ1}(95%CI)=1.21(0.84; 1.75)), but similar ORs were found using Adebamowo's (OR_{Q4vsQ1}(95%CI)=1.49(1.05; 2.12)) and Terry's (OR_{Q4vsQ1}(95%CI)=1.66(1.18; 2.35)) scores.

No association was found between a high compliance with the Castelló's *Prudent* pattern and BC risk (OR_{Q4vsQ1}(95%CI)=1.03(0.75; 1.41)). This absence of association was also observed for Adebamowo's *Prudent* (OR_{Q4vsQ1}(95%CI)=0.77 (0.56; 1.05)) and Terry's *Healthy* (OR_{Q4vsQ1}(95%CI)=0.81(0.59; 1.10)). The ORs under 1 and closer to significance for the case of Adebamowo's *Prudent* pattern are also in agreement with its greater congruence with Castelló's *Mediterranean* (CC_{med}=0.88) than with Terry's *Healthy* (CC=0.77). Bessaoud's *Mediterranean* was the pattern with the highest congruence with Castello's *Mediterranean,* which is reflected in the similarity of the associations with BC found for these two patterns (OR_{Q4vsQ1}(95%CI)=0.72 (0.51; 1.02) and OR_{Q4vsQ1}(95%CI)=0.50 (0.35; 0.71) respectively).

Comparison of CC and Corr as pattern similarity meassurement tools

Despite the fact that all correlations were statistically significant, only when the CC between pattern loadings were ≥0.82 or correlations between pattern scores were ≥0.57, patterns appeared to have a very similar composition and were similarly associated with BC. The same direction of the associations but loss of significance was observed for values of the CC between pattern loadings ≤0.77 and values of the correlation between pattern scores ≤0.52.

DISCUSSION

A high congruence between Castelló's Western pattern and Adebamowo's and Terry's counterpart; between Castello's and Bessaoud's Mediterranean; and between Castelló's Prudent with Adebamowo's Prudent and Terry's Healthy was found in terms of food composition and association with BC risk, independently of the different loading assigned to each food group. The application of dietary patterns from the three selected studies to the EpiGEICAM sample was possible because the authors of these studies provided sufficient detail of the food groupings and of their associated pattern loadings. CC between loadings should be used to assess pattern similarity, instead of relying exclusively on the significance of the Corr between adherence scores.

Numerous nutritional epidemiologists argue that focusing on overall dietary patterns rather than individual foods or nutrients may better capture dietary variability in the population's diet while allowing the evaluation of interactions between dietary factors [4-6]. However, some limitations of this approach have also been identified [4-6, 11, 12, 26, 41]. One of the main criticisms is the potential for subjective interpretations by the investigator to be introduced at various stages of the dietary patterns' construction. Subjective decisions that might affect the comparability between studies are: which foods should be included in each of the defined groups, the thresholds chosen to determine the contribution of food groups to the identified dietary patterns, and the assignation of a label to each of these patterns. However, the present results demonstrate that such limitations can be overcome by a detailed analysis, at least when comprehensive information on food grouping and loadings is provided by authors. The results from four studies were compared taking into account the composition of food groups and patterns to evaluate similarities and differences among them. The conclusions extracted from this comparison were very congruent with the conclusions drawn from the analysis of the association between such patterns and BC risk, demonstrating that comparison is possible by performing a careful analysis of the situation.

Another major concern about data-driven dietary patterns is their applicability to different populations, which can certainly be an issue when comparing different cultures. Even in the case of very population-specific dietary patterns (such as the *Mediterranean* pattern) that are more difficult to identify in some settings (such as northern European countries), the application of these patterns is possible as far as similar food groupings are feasible. The inter-correlation between foods that determines the original structure of patterns might not be reproduced in independent populations, but this does not limit their applicability in such settings. Furthermore, if one pattern has been related to disease in one population, it might be interesting to confirm such an association in an independent population, even if the correlation between foods is not as high as it was in the original study. This is, in fact, the basis of investigator-driven defined patterns, widely applied in different populations to associate them with the occurrence of diverse diseases [1-3, 8, 9, 42]. In a similar way, data-driven dietary patterns also result in a score and, therefore, can and should be replicated in independent populations without methodological questioning.

Schulze et al. [13] have already demonstrated that simplified dietary patterns can be successful for constructing less data-dependent pattern variables that are applicable to populations different to the one from which they have been extracted. This overcomes one of the most important limitations of this methodology and allows the comparison of results across studies. However, Schulze's approach assumes that food groups with a high contribution to one pattern have similar high-loadings and exclude those with lower loadings. This assumption could be relaxed by weighting the sum in the simplified patterns, making this methodology more widely applicable and less dependent on the pattern loadings' variability. Therefore, it is essential to report a detailed composition of food groups, and their loadings resulting from PCA or FA to allow for replication without restrictions.

As explained in the introduction, the validity and reproducibility of investigator-driven dietary patterns has been explored [1-3, 8, 9] within the Dietary Patterns Methods Project [43]. With regard to data-driven dietary patterns, various studies have assessed their validity by comparing patterns extracted in the same population using information obtained with different assessment tools (FFQ vs 24 hour recall) [18, 19, 32, 33, 35] or applying different statistical approaches [28, 34]. Some have also assessed their reproducibility by comparing dietary patterns extracted in the same population with dietary information obtained with common assessment tools in different moments of time [17-20]. However, to our knowledge this is the first study assessing the applicability of data-driven dietary patterns to a population different from the one that originated them, and the first to use CC to determine pattern similarity. To establish conclusive evidence regarding associations between dietary patterns and disease, similar results need to be obtained in different populations. Although the comparison of independently developed data-driven dietary patterns and their association with disease is valid to establish evidence of associations, the application of the same dietary patterns in different populations is also necessary. This should overcome some of the aforementioned limitations of dietary pattern analysis.

Finally, these results are in agreement with the threshold that various authors have set for the CC [14-16], indicating that a value in the range [0.85-0.94] results in fair similarity between components (dietary patterns in this case) and a value ≥0.95 implies equivalent composition [14, 15]. In the present study, a similar direction, magnitude and significance of the association for values of the CC between [0.86-0.95] (Corr between [0.67-0.85]) and a loss of the significance of the original associations in the applied patterns when CC ranged between [0.77-0.82] (Corr between [0.52-0.57]) was observed. All correlations were statistically significant but only Corr≥0.67 correspond with CC≥0.85 and with similar associations between the compared patterns and BC risk. These results indicate, for the first time, that significance of correlations between pattern scores is not sufficient to ascertain pattern similarity, showing that the CC could be a more appropriate measure for evaluating such similarity.

CONCLUSION

The current results indicate that applying data-driven dietary patterns in different settings from the one from which they were extracted is possible independently from the labelling used by authors, provided that they come from similar populations and patterns composition is interpreted cautiously. The publication of information on food grouping, pattern composition and loadings is essential to allow for replication. The congruence coefficient between pattern loadings should be used to evaluate similarity between patterns, rather than relying solely on the statistical significance of simple linear correlations between pattern scores.

Table 1: Castelló's [21], Bessaoud's [28], Adebamowo's [29] and Terry's [30] food groups and pattern loadings extracted from the original publications[21,28,29,30] omitting the loadings whose values are under |0.15|.

Castelló et al.				Bessaoud et al.			Adebamowo et al.			Terry et al.		
Group Name	**West^{a}**	**Prud^{b}**	**Med^{c}**	**Group Name**	**West^{a}**	**Med^{c}**	**Group Name**	**West^{a}**	**Prud^{c}**	**Group Name**	**West^{a}**	**Heal^{d}**
High-fat dairy	0.60	0.00^{e}	0.20	Cheese	0.35	0.00^{e}	High-fat dairy	0.31	0.00^{e}	High-fat dairy	0.46	0.00^{e}
Low-fat dairy	-0.49	0.60	0.00^{e}	Dairy products	0.16	0.00^{e}	Low-fat dairy	0.00^{e}	0.32	Low-fat dairy	0.00^{e}	0.40
Eggs	0.19	0.00^{e}	0.16	Eggs	0.45	0.00^{e}	Eggs	0.36	0.00^{e}	Eggs	0.21	0.32
White meat	0.00^{e}	0.17	0.18	Poultry	0.26	0.18	Poultry	0.19	0.31	Poultry	0.00^{e}	0.36
Red meat	0.27	0.00^{e}	0.22	Meat	0.00^{e}	0.00^{e}	Red meat	0.61	0.00^{e}	Meat	0.46	0.33
				Offal and giblets	0.00^{e}	0.18						
				Hamburger	0.28	0.00^{e}						
Proc. meat	0.36	0.00^{e}	0.26	Proc. meats	0.46	0.00^{e}	Proc. meat	0.56	0.00^{e}	Proc. meat	0.58	0.00^{e}
White fish	0.00^{e}	0.22	0.34	Lean fish	0.00^{e}	0.48	Fish	0.00^{e}	0.42	Fish	0.00^{e}	0.54
Oily fish	0.00^{e}	0.24	0.44	Fatty fish	0.00^{e}	0.52		0.00^{e}				
Shellfish	0.17	0.27	0.35	Mollusk and shell.	0.00^{e}	0.30		0.00^{e}				
Leafy vegetables	0.00^{e}	0.34	0.40	Raw veg	0.00^{e}	0.63	Leafy veg	0.00^{e}	0.65	Vegetables	0.00^{e}	0.66
Fruiting vegetables	0.00^{e}	0.36	0.45	Cooked veg	0.00^{e}	0.63	Tomatoes	0.00^{e}	0.54			
Root vegetables	0.00^{e}	0.35	0.44				Dark yellow veg	0.00^{e}	0.62			
Other vegetables	0.00^{e}	0.40	0.42				Other veg	0.00^{e}	0.69			
							Cruciferous veg	0.00^{e}	0.60			
							Onions	0.00^{e}	0.48			
							Garlic	0.00^{e}	0.32			
Legumes	0.21	0.15	0.34	Legumes	0.32	0.33	Legumes	0.00^{e}	0.61			
Potatoes	0.17	0.25	0.40				Potatoes	0.37	0.26	Potato	0.43	0.00^{e}
Fruits	0.00^{e}	0.31	0.31	Fruits	0.16	0.42	Fruit	0.00^{e}	0.63	Fruit	0.00^{e}	0.55

Nuts	0.18	0.22	0.29				Nuts	0.28	0.00[e]			
Refined grains	0.37	0.15	0.23	Cereals	0.43	0.19	Refined grains	0.64	0.19	Refined grains	0.54	0.00[e]
Whole grains	-0.43	0.47	0.00[e]				Whole grains	0.00[e]	0.45	Whole grains Cereal	0.20 0.00[e]	0.43 0.34
Olives and veg. oil	0.00[e]	0.19	0.34	Olive oil Other oil	0.00[e] 0.42	0.69 0.00[e]						
Other edible fats	0.22	0.00[e]	0.00[e]	Butter	0.43	0.00[e]	Margarine Butter	0.37 0.19	0.00[e]	Margarine	0.00[e]	0.26
Sweets	0.35	0.18	0.00[e]	Sweets	0.61	0.00[e]	Desserts	0.57	0.00[e]	Sweets	0.54	-0.17
Sugary	0.24	0.00[e]	0.00[e]									
Juices	0.25	0.67	-0.39				Fruit juice	0.00[e]	0.30	Juice	0.00[e]	0.27
Caloric drinks	0.74	0.21	-0.25				High-sugar drinks	0.36	0.00[e]	Soda	0.45	0.00[e]
Conv food & sauces	0.47	0.00[e]	0.24	Pizzas	0.45	0.00[e]	Salad dressing French fries Pizza Snacks Mayonnaise Condiments	0.00[e] 0.55 0.46 0.44 0.31 0.21	0.41 0.00[e] 0.00[e] 0.17 0.00[e] 0.00[e]	Snacks	0.16	0.00[e]

[a] Western Pattern; [b] Prudent Pattern; [c] Mediterranean Pattern; [d] Healthy

[e] Since Adebamowo et al. and Terry et al. only showed component loadings >=|0.15|, we assign the value 0.00 to component loadings <|0.15| in all studies.

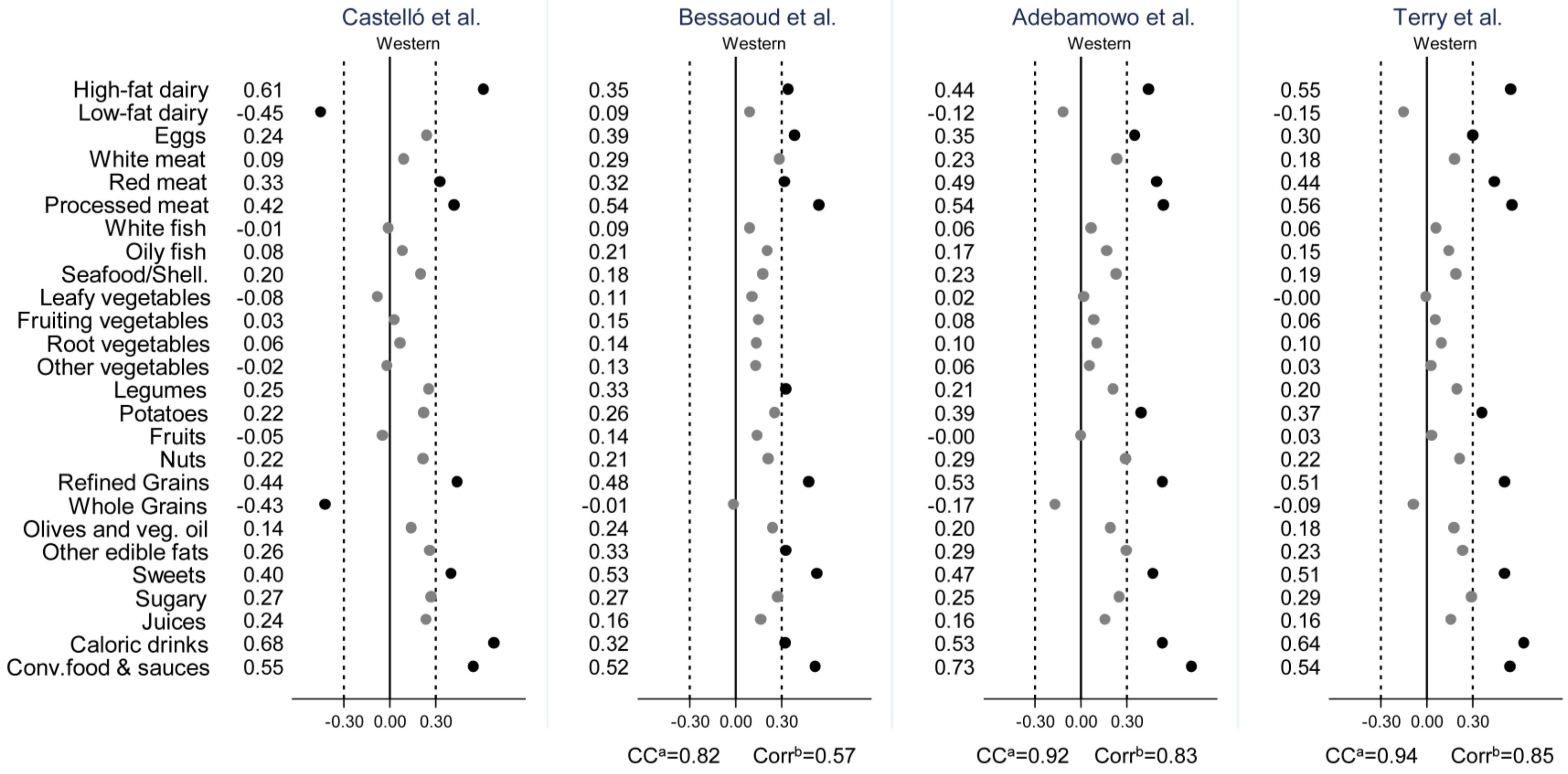
Table 2: Adjusted OR of breast cancer per quartiles and standard deviation increase in the adherence to Castelló's [21], Bessaoud's [28], Adebamowo's [29] and Terry's [30] recalculated dietary patterns.

	Castelló et al.		**Bessaoud et al.**		**Adebamowo et al.**		**Terry et al.**	
Western	**Controls /Cases**	**OR[a](95%CI)**	**Controls /Cases**	**OR[a](95%CI)**	**Controls /Cases**	**OR[a](95%CI)**	**Controls /Cases**	**OR[a](95%CI)**
Quartiles								
Q1	244/192	1.00	244/214	1.00	243/198	1.00	244/180	1.00
Q2	242/231	1.25 (0.94;1.67)	243/206	0.91 (0.68;1.23)	244/232	1.15 (0.86;1.52)	243/246	1.40 (1.06;1.85)
Q3	244/254	1.30 (0.96;1.74)	242/272	1.24 (0.91;1.70)	243/239	1.13 (0.84;1.53)	242/245	1.32 (0.96;1.80)
Q4	243/296	1.50 (1.09;2.07)	244/281	1.21 (0.84;1.75)	243/304	1.49 (1.05;2.12)	244/302	1.66 (1.18;2.35)
Per increase in 1 SD		1.17 (1.04;1.31)		1.05 (0.92;1.21)		1.13 (0.98;1.29)		1.13 (0.99;1.28)
Prudent/Healthy	**Controls /Cases**	**OR[a](95%CI)**	**Controls /Cases**	**OR[a](95%CI)**	**Controls /Cases**	**OR[a](95%CI)**	**Controls /Cases**	**OR[a](95%CI)**
Quartiles								
Q1	244/228	1.00			244/242	1.00	243/255	1.00
Q2	243/244	1.08 (0.83;1.42)			243/250	1.00 (0.76;1.31)	244/232	0.90 (0.69;1.19)
Q3	243/229	1.03 (0.77;1.38)			243/249	0.97 (0.73;1.30)	243/226	0.83 (0.62;1.10)
Q4	243/272	1.03 (0.75;1.41)			243/232	0.77 (0.56;1.05)	243/260	0.81 (0.59;1.10)
Per increase in 1 SD		1.00 (0.89;1.13)				0.89 (0.79;1.00)		0.94 (0.84;1.05)
Mediterranean	**Controls /Cases**	**OR[a](95%CI)**	**Controls /Cases**	**OR[a](95%CI)**	**Controls /Cases**	**OR[a](95%CI)**	**Controls /Cases**	**OR[a](95%CI)**
Quartiles								
Q1	243/262	1.00	244/251	1.00				
Q2	244/247	0.90 (0.69;1.18)	243/241	0.94 (0.70;1.25)				
Q3	242/267	0.83 (0.61;1.12)	243/244	0.90 (0.66;1.22)				
Q4	244/197	0.50 (0.35;0.71)	243/237	0.72 (0.51;1.02)				
Per increase in 1 SD		0.78 (0.68;0.88)		0.88 (0.77;1.00)				

[a] Adjusted by total energy intake, alcohol consumption, body mass index (BMI) from self-reported weight and height (BMI=Kg/m^2), physical activity in the last year, smoking, education, history of breast disease other than cancer, family history of BC, age at menarche, age at first delivery and menopausal status

FIGURE LEGENDS

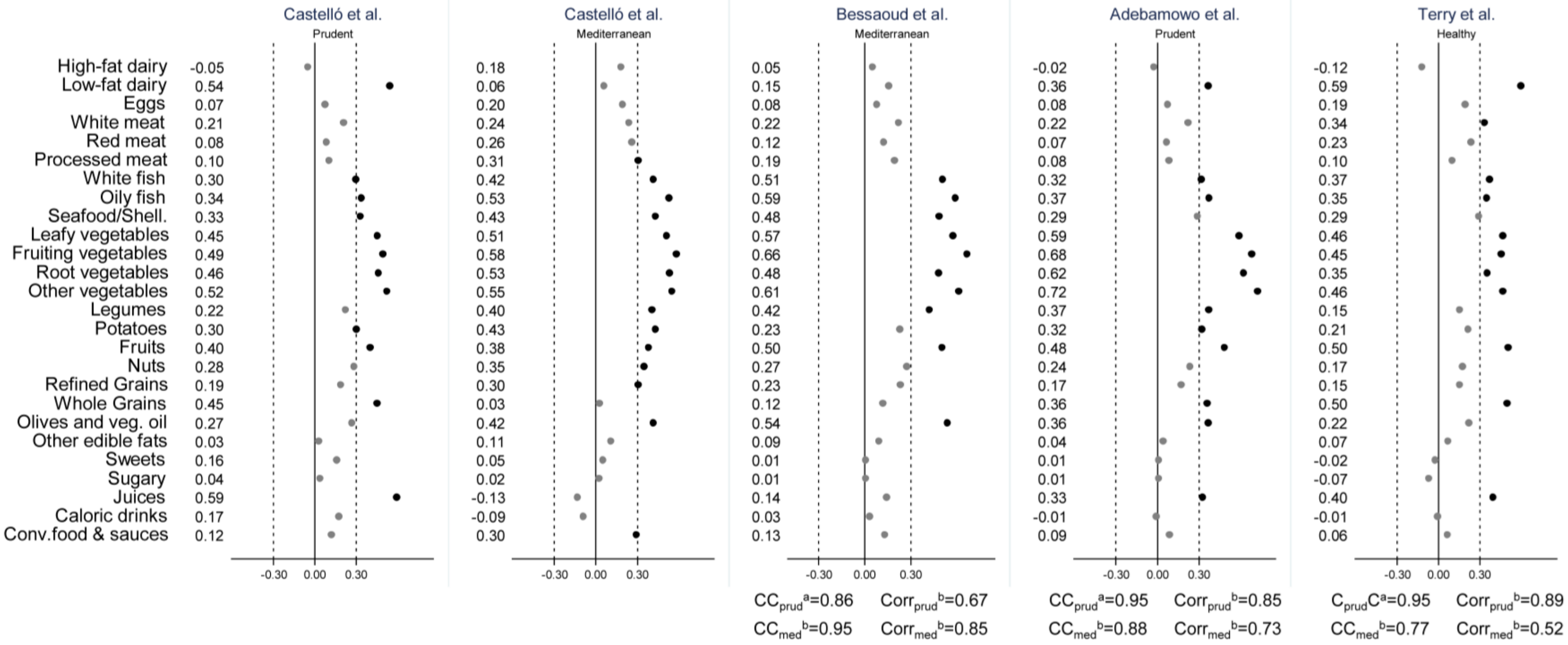
Figure 1: Linear correlation (pattern loadings) between food consumption and Castelló's [21], Bessaoud's [28], Adebamowo's [29] and Terry's [30] Western pattern scores. Congruence coefficients between component loadings and correlation coefficients between component scores of Castelló et al.[21] *Western* pattern with Bessaoud's [28], Adebamowo's [29] and Terry's [30] Western *pattern*.



[a] Congruence coefficients for agreement between component loadings of Castello et al. with Bessaoud, Adebamowo and Terry.

[b] Correlation coefficients for agreement between component scores of Castello's et al with Bessaoud's, Adebamowo's and Terry's component scores. All correlations were statistically sifnificant at a 95% confidence level.

Figure 2: Linear correlation (pattern loadings) between food consumption and Castelló's [21] Prudent, Castelló's [21] Mediterranean, Bessaoud's [28] Mediterranean, Adebamowo's [29] Prudent and Terry's [30] Healthy patterns. Congruence coefficients between component loadings and correlation coefficients between component scores of Castelló et al. [21] Prudent and Mediterranean patterns with Bessaoud's [28] Mediterranean, Adebamowo's [29] Prudent and Terry's [30] Healthy patterns



[a] Congruence coefficients for agreement between component loadings of Castello et al. with Bessaoud, Adebamowo and Terry.

[b] Correlation coefficients for agreement between component scores of Castello's et al with Bessaoud's, Adebamowo's and Terry's component scores. All correlations were statistically sifnificant at a 95% confidence level.

REFERENCES

1. Harmon BE, Boushey CJ, Shvetsov YB, et al. Associations of key diet-quality indexes with mortality in the Multiethnic Cohort: the Dietary Patterns Methods Project. *Am J Clin Nutr.* 2015;101:587-597.

2. McCullough ML. Diet patterns and mortality: common threads and consistent results. *The Journal of nutrition.* 2014;144:795-796.

3. Reedy J, Krebs-Smith SM, Miller PE, et al. Higher diet quality is associated with decreased risk of all-cause, cardiovascular disease, and cancer mortality among older adults. *The Journal of nutrition.* 2014;144:881-889.

4. Barkoukis H. Importance of understanding food consumption patterns. *J Am Diet Assoc.* 2007;107:234-236.

5. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol.* 2002;13:3-9.

6. Jacques PF, Tucker KL. Are dietary patterns useful for understanding the role of diet in chronic disease? *Am J Clin Nutr.* 2001;73:1-2.

7. Fung TT, McCullough ML, Newby PK, et al. Diet-quality scores and plasma concentrations of markers of inflammation and endothelial dysfunction. *Am J Clin Nutr.* 2005;82:163-173.

8. Liese AD, Krebs-Smith SM, Subar AF, et al. The Dietary Patterns Methods Project: synthesis of findings across cohorts and relevance to dietary guidance. *The Journal of nutrition.* 2015;145:393-402.

9. George SM, Ballard-Barbash R, Manson JE, et al. Comparing indices of diet quality with chronic disease mortality risk in postmenopausal women in the Women's Health Initiative Observational Study: evidence to inform national dietary guidance. *Am J Epidemiol.* 2014;180:616-625.

10. Gleason PM, Boushey CJ, Harris JE, Zoellner J. Publishing nutrition research: a review of multivariate techniques--part 3: data reduction methods. *Journal of the Academy of Nutrition and Dietetics.* 2015;115:1072-1082.

11. Martinez ME, Marshall JR, Sechrest L. Invited commentary: Factor analysis and the search for objectivity. *Am J Epidemiol.* 1998;148:17-19.

12. Slattery ML, Boucher KM. The senior authors' response: Factor analysis as a tool for evaluating eating patterns. *Am J Epidemiol.* 1998;148:20-21.

13. Schulze MB, Hoffmann K, Kroke A, Boeing H. An approach to construct simplified measures of dietary patterns from exploratory factor analysis. *The British journal of nutrition.* 2003;89:409-419.

14. Haven S, Berge J. Tucker’s coefficient congruence as a measure of factorial invariance: An empirical study. *Heymans Bulletin 290 EX. University of Groningen.* 1977.

15. Lorenzo-Seva U, Berge J. Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology.* 2006;2:54-67.

16. Nesselroade J, Baltes P. On a dilemma of comparative factor analysis: A study of factor matching based on random data. *Educational and Psychological Measurement.* 1970;30:935-948.

17. Hu FB, Rimm E, Smith-Warner SA, et al. Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *Am J Clin Nutr.* 1999;69:243-249.

18. Khani BR, Ye W, Terry P, Wolk A. Reproducibility and validity of major dietary patterns among Swedish women assessed with a food-frequency questionnaire. *The Journal of nutrition.* 2004;134:1541-1545.

19. Nanri A, Shimazu T, Ishihara J, et al. Reproducibility and validity of dietary patterns assessed by a food frequency questionnaire used in the 5-year follow-up survey of the Japan Public Health Center-Based Prospective Study. *Journal of epidemiology / Japan Epidemiological Association.* 2012;22:205-215.

20. Newby PK, Weismayer C, Akesson A, Tucker KL, Wolk A. Long-term stability of food patterns identified by use of factor analysis among Swedish women. *The Journal of nutrition.* 2006;136:626-633.

21. Castello A, Pollan M, Buijsse B, et al. Spanish Mediterranean diet and other dietary patterns and breast cancer risk: case-control EpiGEICAM study. *British journal of cancer.* 2014;111:1454-1462.

22. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol.* 1985;122:51-65.

23. Vioque J, Weinbrenner T, Asensio L, Castello A, Young IS, Fletcher A. Plasma concentrations of carotenoids and vitamin C are better correlated with dietary intake in normal weight than overweight and obese elderly subjects. *The British journal of nutrition.* 2007;97:977-986.

24. Vioque J, Navarrete-Munoz EM, Gimenez-Monzo D, et al. Reproducibility and validity of a food frequency questionnaire among pregnant women in a Mediterranean area. *Nutrition journal.* 2013;12:26.

25. Albuquerque RC, Baltar VT, Marchioni DM. Breast cancer and dietary patterns: a systematic review. *Nutrition reviews.* 2014;72:1-17.

26. Brennan SF, Cantwell MM, Cardwell CR, Velentzis LS, Woodside JV. Dietary patterns and breast cancer risk: a systematic review and meta-analysis. *Am J Clin Nutr.* 2010;91:1294-1302.

27. Edefonti V, Randi G, La Vecchia C, Ferraroni M, Decarli A. Dietary patterns and breast cancer: a review with focus on methodological issues. *Nutrition reviews.* 2009;67:297-314.

28. Bessaoud F, Tretarre B, Daures JP, Gerber M. Identification of dietary patterns using two statistical approaches and their association with breast cancer risk: a case-control study in Southern France. *Ann Epidemiol.* 2012;22:499-510.

29. Adebamowo CA, Hu FB, Cho E, Spiegelman D, Holmes MD, Willett WC. Dietary patterns and the risk of breast cancer. *Ann Epidemiol.* 2005;15:789-795.

30. Terry P, Suzuki R, Hu FB, Wolk A. A prospective study of major dietary patterns and the risk of breast cancer. *Cancer Epidemiol Biomarkers Prev.* 2001;10:1281-1285.

31. Burt C. Factor Analysis and canonical correlations. *British Journal of Statistical Psychology.* 1948;1:95-106.

32. Ambrosini GL, O'Sullivan TA, de Klerk NH, Mori TA, Beilin LJ, Oddy WH. Relative validity of adolescent dietary patterns: a comparison of a FFQ and 3 d food record. *The British journal of nutrition.* 2011;105:625-633.

33. Appannah G, Pot GK, O'Sullivan TA, Oddy WH, Jebb SA, Ambrosini GL. The reliability of an adolescent dietary pattern identified using reduced-rank regression: comparison of a FFQ and 3 d food record. *The British journal of nutrition.* 2014;112:609-615.

34. Castro MA, Baltar VT, Selem SS, Marchioni DM, Fisberg RM. Empirically derived dietary patterns: interpretability and construct validity according to different factor rotation methods. *Cadernos de saude publica.* 2015;31:298-310.

35. Okubo H, Murakami K, Sasaki S, et al. Relative validity of dietary patterns derived from a self-administered diet history questionnaire using factor analysis among Japanese adults. *Public health nutrition.* 2010;13:1080-1089.
36. Rencher A. Principal Component Analysis. *Methods of Multivariate Analysis.* : John Wiley & Sons, Inc.; 2002:380-407.
37. Tucker LR. A method for the synthesis of factor analysis studies (Personnel Research Section Report No. 984). Washington: Department of the Army. 1951:18-20.
38. Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *Am J Epidemiol.* 2010;171:624-632.
39. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in medicine.* 2011;30:377-399.
40. Royston P, White IR. Multiple Imputation By Chained Equiations (MICE): Implementation in Stata. *Journal of Statistical Software.* 2011;45:1-19.
41. Michels KB, Schulze MB. Can dietary patterns help us detect diet-disease associations? *Nutrition research reviews.* 2005;18:241-248.
42. Fung TT, Hu FB, McCullough ML, Newby PK, Willett WC, Holmes MD. Diet quality is associated with the risk of estrogen receptor-negative breast cancer in postmenopausal women. *The Journal of nutrition.* 2006;136:466-472.
43. National Cancer Institute. Division of Cancer Control & Population. Dietary Patterns Methods Project. http://epi.grants.cancer.gov/dietary-patterns/. Revised August 3,2015. Accesed March 2, 2016.

Conflict of Interest Disclosure

The authors declare that they have no conflict of interest.

SUPPLEMENTARY MATERIAL

Supplemental Table 1: Description of Castelló et al., Bessaoud et al., Adebamowo et al. and Terry et al. study characteristics and main results.

Authors	**Study design**	**Country**	**n**	**Breast cancer cases**	**Participants' Age (Years)**	**Patterns**	**Association**
Castelló et al. (2014)[21]	Case-control	Spain	1946	973	22-71	Western	OR[a](95%CI)=1.17 (1.04–1.31)
						Prudent	OR[a](95%CI)=1.00 (0.89–1.13)
						Mediterraean	OR[a](95%CI)=0.78 (0.69–0.89)
Bessaoud et al. (2012)[28]	Case-control	France	1359	437	25-85	Western	OR[a](95%CI)=0.88(0.73;1.06)
						Mediterranean	OR[a](95%CI)=1.08 (0.93;1.25)
						Meat-eaters and drinkers	OR[a](95%CI)=1.20(1.04;1.38)
Adebamowo et al. (2005)[29]	Prospective cohort	USA	90638	710	30-50	Western	RR[b](95%CI)=0.97(0.71;1.33)
						Prudent	RR[b](95%CI)=0.90 (0.68;1.18)
Terry et al. (2001)[30]	Prospective cohort	Sweden	61463	1328	40-76	Western	RR[b](95%CI)=1.00(0.79;1.26)
						Healthy	RR[b](95%CI)=0.92 (0.76;1.13)
						Drinker	RR[b](95%CI)=1.27 (1.06-1.52)

[a] OR of breast cancer according to an increment of one standard error in the score of adherence for each pattern

[b] Multivariate RR comparing highest to lowest quintiles of cumulative average score

Supplemental Table 2: Composition of food groups from Castelló's[21], Bessaoud's[28], Adebamowo's[29] and Terry's[30] studies[a]

Castelló et al.	Bessaoud et al.	Adebamowo et al.	Terry et al.
HIGH-FAT DAIRY: whole-fat milk; cream; condensed milk; whole-fat yogurt; high-fat cheese; custard, flan, pudding; ice-cream.	DAIRY PRODUCTS: whole-fat milk; cream; condensed milk; whole-fat yogurt; custard, flan, pudding; ice-cream; low-fat milk; low-fat yogurt; CHEESE: high-fat cheese; cottage or fresh white cheese.	HIGH-FAT DAIRY: whole-fat milk; cream; condensed milk; whole-fat yogurt; high-fat cheese; custard, flan, pudding; ice-cream.	HIGH-FAT DAIRY: whole-fat milk; cream; condensed milk; whole-fat yogurt; high-fat cheese; custard, flan, pudding; ice-cream.
LOW-FAT DAIRY: low-fat milk; low-fat yogurt; cottage or fresh white cheese.		LOW-FAT DAIRY: low-fat milk; low-fat yogurt; cottage or fresh white cheese.	LOW-FAT DAIRY: low-fat milk; low-fat yogurt; cottage or fresh white cheese.
EGGS: eggs	EGGS: eggs	EGGS: eggs	EGGS: eggs
WHITE MEAT: chicken; game (turkey, rabbit. etc.)	POULTRY: chicken; game (turkey, rabbit. etc.)	POULTRY: chicken; game (turkey, rabbit. etc.)	POULTRY: chicken; game (turkey, rabbit. etc.)
RED MEAT: pork; beef; lamb; liver; intestines, brains and sweetbreads; hamburger.	MEAT: pork; beef; lamb.	RED MEAT: pork; beef; lamb; liver; intestines, brains and sweetbreads; hamburger.	MEAT: pork; beef; lamb; intestines, brains and sweetbreads; hamburger.
	OFFALS AND GIBLETS: liver; intestines, brains and sweetbreads.		
	HAMBURGER: hamburger.		
PROC. MEAT: cold meat; sausages; bacon; pâté, foie-gras	PROC. MEATS: cold meat; sausages; bacon; pâté, foie-gras	PROC. MEAT: cold meat; sausages; bacon; pâté, foie-gras	PROCESSED MEAT: cold meat; sausages; bacon; pâté, foie-gras
WHITE FISH: fresh white fish: hake, sea bass, sea bream;	LEAN FISH: fresh white fish: hake, sea bass, sea bream;	FISH: fresh white fish: hake, sea bass, sea bream; fresh big blue fish: tuna, swordfish; other fresh blue fish: sardines, anchovies, salmon; canned tuna canned sardines or mackerel; salted and smoked fish; clams, mussels, oysters, squid, cuttlefish, octopus, crustaceans: prawn, crab, shrimp, lobster	FISH: fresh white fish: hake, sea bass, sea bream; fresh big blue fish: tuna, swordfish; other fresh blue fish: sardines, anchovies, salmon; canned tuna canned sardines or mackerel; salted and smoked fish; clams, mussels, oysters, squid, cuttlefish, octopus, crustaceans: prawn, crab, shrimp, lobster
OILY FISH: fresh big blue fish: tuna, swordfish; other fresh blue fish: sardines, anchovies, salmon; canned tuna canned sardines or mackerel; salted and smoked fish	FATTY FISH: fresh big blue fish: tuna, swordfish; other fresh blue fish: sardines, anchovies, salmon; canned tuna canned sardines or mackerel; salted and smoked fish		
SHELLFISH: clams, mussels, oysters, squid, cuttlefish, octopus, crustaceans: prawn, crab, shrimp, lobster	MOLLUSK AND SHELL: clams, mussels, oysters, squid, cuttlefish, octopus, crustaceans: prawn, crab, shrimp, lobster		

LEAFY VEGETABLES: spinach or chard; lettuce, endive, escarole. FRUITING VEGETABLES: tomato; eggplant, zucchini and cucumber; pepper; artichoke. ROOT VEGETABLES: carrot, pumpkin. OTHER VEGETABLES: cooked cabbage, cauliflower, broccoli; onion; green beans, asparagus; corn; garlic	RAW VEG[b]: lettuce, endive, escarole; tomato; onion*0.25; (carrot, pumpkin)*0.25;(eggplant, zucchini and cucumber)*0.33; garlic*0.25; COOKED VEG[b]: spinach orchard; onion*0.75; (carrot, pumpkin)*0.25; (eggplant, zucchini and cucumber)*0.67; pepper; artichoke; cooked cabbage, cauliflower, broccoli;; green beans, asparagus; corn; garlic*0.75	LEAFY VEG: spinach or chard; lettuce, endive, escarole. TOMATOES: tomato DARK YELLOW VEG: carrot, pumpkin. OTHER VEG: eggplant, zucchini and cucumber; pepper; artichoke; green beans, asparagus; corn; CRUCIFEROUS VEG: cooked cabbage, cauliflower, broccoli ONIONS: onion. GARLIC: garlic.	VEGETABLES: spinach or chard; lettuce, endive, escarole; tomato; eggplant, zucchini and cucumber; pepper; artichoke; carrot, pumpkin, cooked cabbage, cauliflower, broccoli; onion; green beans, asparagus; corn; garlic.
LEGUMES: legumes	LEGUMES: legumes	LEGUMES: legumes	
POTATOES: roasted or boiled potatoes.		POTATOES: roasted or boiled potatoes.	POTATO: roasted or boiled potatoes; french fries.
FRUITS: orange, mandarin, banana; apple, pear; peach, nectarine, apricot; watermelon, melon; grapes; plums, prunes (dried or fresh); kiwi.	FRUITS: orange, mandarin, banana; apple, pear; peach, nectarine, apricot; watermelon, melon; grapes; plums, prunes (dried or fresh); kiwi.	FRUIT: orange, mandarin, banana; apple, pear; peach, nectarine, apricot; watermelon, melon; grapes; plums, prunes (dried or fresh); kiwi.	FRUIT: orange, mandarin, banana; apple, pear; peach, nectarine, apricot; watermelon, melon; grapes; plums, prunes (dried or fresh); kiwi.
NUTS: almonds, peanuts, hazelnuts.		NUTS: almonds, peanuts, hazelnuts.	
REFINED GRAINS: white-flour bread; rice; pasta.	CEREALS: white-flour bread; rice; pasta; whole-grain bread and partial whole-grain bread; breakfast cereals.	REFINED GRAINS: white-flour bread; rice; pasta.	REFINED GRAINS: white-flour bread; rice; pasta.
WHOLE GRAINS: whole-grain bread and partial whole-grain bread; breakfast cereals.		WHOLE GRAINS: whole-grain bread and partial whole-grain bread; breakfast cereals.	WHOLE GRAINS: whole-grain bread and partial whole-grain bread; CEREAL: breakfast cereals.
OLIVES AND VEG. OIL: Olives; Added olive oil to salads, bread and dishes; Other vegetable oils: sunflower, corn, soybean.	OLIVE OIL: Added olive oil to salads, bread and dishes; OTHER OIL: Other vegetable oils: sunflower, corn, soybean.		
OTHER EDIBLE FATS:	BUTTER: butter	MARGARINE: margarine	MARGARINE: margarine; butter

margarine; butter.		BUTTER: butter	
SWEETS: chocolate, sweets and similar; cocoa powder and similar; plain cookies; chocolate cookies; pastries: croissant, donut, cake, pie;	SWEETS: chocolate, sweets and similar; cocoa powder and similar; plain cookies; chocolate cookies; pastries: croissant, donut, cake, pie; jam, honey; sugar	DESSERTS: chocolate, sweets and similar; cocoa powder and similar; plain cookies; chocolate cookies; pastries: croissant, donut, cake, pie: jam, honey; sugar	SWEETS: chocolate, sweets and similar; cocoa powder and similar; plain cookies; chocolate cookies; pastries: croissant, donut, cake, pie; jam, honey; sugar
SUGARY: jam, honey; sugar			
JUICES: freshly squeezed orange juice; non freshly squeezed juice		FRUIT JUICE: freshly squeezed orange juice; non freshly squeezed juice	JUICE: freshly squeezed orange juice; non freshly squeezed juice
CALORIC DRINKS: sugar-sweetened soft drinks.		HIGH-SUGAR DRINKS: sugar-sweetened soft drinks.	SODA: sugar-sweetened soft drinks.
CONV FOOD & SAUCES :fish sticks; french fries; chips; pizza; croquettes; mayonnaise; tomato sauce; ketchup	PIZZAS: pizza	SALAD DRESSING: Olives; Added olive oil to salads, bread and dishes; Other vegetable oils: sunflower, corn, soybean. FRENCH FRIES: french fries. PIZZA: pizza SNACKS: chips MAYONNAISE: mayonnaise. CONDIMENTS: tomato sauce; ketchup	SNACKS: chips

[a] Separated by “,” foods whose consumption is collected jointly and separated by “;”foods whose consumption is collected separately.

[b] The questionnaire from the present study did not collect whether the vegetables were consumed cooked or raw. We distributed them across categories by weighting the intake according to the common Spanish habits.

Supplementary Example 1: Explanation of the calculations carried out in steps 1-3 of the "Applicability" subsection of the "Methods" section.

1. Food consumption (in grams) collected within EpiGEICAM study was grouped into the food groups defined by Castelló et al. and Bessaoud et al. as described in Supplementary Table 2.

2. Pattern scores of adherence to the Castelló's *Western* Pattern and to the Bessaoud's *Western* pattern were calculated for each women (i=1,…1946) as the linear combination of their food group consumption (constructed in step 1), weighted by the original pattern loadings reported by these studies and summarized in **Table 1** of the manuscript.

The score for Castelló et al. *Western* pattern for women i ($WSC_i;\ i=1,\ldots1946$) was calculated as follows:

$$WSC_i= \text{High-fat dairy}_i*0.60+ \text{Low fat dairy}_i*-0.49 + \text{Eggs}_i*0.19 + \text{White meat}_i*0.00 + \text{Red meat}_i*0.27 + \text{Proc. Meat}_i*0.36 + \text{White fish}_i*0.00 + \text{Oily fish}_i*0.00 + \text{Shellfish}_i*0.17 + \text{Leafy vegetables}_i*0.00+ \text{Fruiting vegetables}_i*0.00 + \text{Root vegetables}_i*0.00 + \text{Other vegetables}_i*0.00+ \text{Legumes}_{i*}0.21 + \text{Potatoes}_i*0.17 + \text{Fruits}_i*0.00 + \text{Nuts}_i*0.18 + \text{Refined grains}_i*0.37 + \text{Whole grains}_i *-0.43 + \text{Olives and veg. Oil}_i*0.00 + \text{Other edible fats}_i*0.22 + \text{Sweets}_i*0.35 + \text{Sugary}_i*0.24 + \text{Juices}_i*0.25 + \text{Caloric drinks}_i*0.74+ \text{Conv food \& sauces}*0.47$$

The score for Bessaoud et al. *Western* pattern for women i ($WSB_i;\ i=1,\ldots1946$) was calculated as follows:

$$WSB_i= \text{Cheese}_i*0.35 + \text{Dairy products}_i*0.16 + \text{Eggs}_i*0.45 + \text{Poultry}_i*0.26 + \text{Meat}_i*0.00 +\text{Offal and giblets}_i*0.00 +\text{Hamburger}_i*0.28 +\text{Proc. meats}_i*0.46 +\text{Lean fish}_i*0.00 +\text{Fatty fish}_i*0.00 +\text{Mollusk and shell.}_i*0.00 + + \text{Raw veg}_i*0.00 + \text{Cooked veg}_i*0.00 +\text{Legumes}_i*0.32 + \text{Fruits}_i*0.16 +\text{Cereals}_i*0.43 +\text{Olive oil}*0.00 + \text{Other oil}*0.42 + \text{Butter}*0.43 + \text{Sweets}*0.61 + \text{Pizzas}*0.45$$

As a first measure to assess the similarity of pairs of patterns, the Pearson's correlation coefficients (Corr) were calculated between the scores of those patterns considered comparable.

Corr= Correlation (WSC_i, WSB_i)=0.57 *(see value in* ***Figure 1*** of the manuscript*)*

3. In order to obtain pattern loadings associated to the same exact food groups, loadings for both *Western* patterns were recalculated using the food definition provided by Castello et al. In agreement with their methodological definition of pattern loadings, they were recalculated by correlating the food group consumption of the 26 groups defined in Castello et al. with the scores calculated in the step 2:

The loadings for Castelló et al. *Western* pattern (LC_j*; j=1,…,26 food groups from Castelló et al.)* summarized in the first column of **Figure 1** of the manuscript were calculated as:

$$LC_j=Corr(F_j, WSC)$$

Where:

F_j*= Each of the i:1,…26 food groups defined in Castelló et al, i.e.: High-fat dairy; Low fat dairy; Eggs; White meat; Red meat; Proc. Meat; White fish; Oily fish; Shellfish; Leafy vegetables; Fruiting vegetables; Root vegetables; Other vegetables; Legumes; Potatoes; Fruits; Nuts; Refined grains; Whole grains ; Olives and veg. Oil; Other edible fats; Sweets; Sugary; Juices; Caloric drinks; Conv food & sauces)*

WSC=Score of adherence to the Western pattern from Castelló et al. calculated in step 2.

The loadings for Bessaoud et al. *Western* pattern (LB_j*; j=1,…,26 food groups from Castelló et al.)* summarized in the second column of **Figure 1** of the manuscript were calculated as:

$$LB_j=Corr(F_j, WSB)$$

Where:

F_j*= Each of the i:1,…26 food groups defined in Castelló et al, i.e.∵: High-fat dairy; Low fat dairy; Eggs; White meat; Red meat; Proc. Meat; White fish; Oily fish; Shellfish; Leafy vegetables; Fruiting vegetables; Root vegetables; Other vegetables; Legumes; Potatoes; Fruits; Nuts; Refined grains; Whole grains ; Olives and veg. Oil; Other edible fats; Sweets; Sugary; Juices; Caloric drinks; Conv food & sauces)*

WSB=Score of adherence to the Western pattern from Bessaoud et al. calculated in step 2.

After obtaining comparable loadings for both *Western* patterns that are associated to the same food groups, the congruence coefficient (CC) was calculated as follows:

$$CC = \frac{\sum_{j=1}^{26} LC_j \cdot LB_j}{\sqrt[2]{\left(\sum_{j=1}^{26} LC_j^2\right) \cdot \left(\sum_{j=1}^{26} LB_j^2\right)}} = 0.82\ (\mathit{See}\ \boldsymbol{Figure}\ \mathbf{1}\ \mathit{of\ the\ manuscript})$$

j==1,…,26 food groups