Reproducibility of data-driven dietary patterns in two groups of adult Spanish women from different studies

Adela Castelló1,2,3*, Virginia Lope1,2,3, Jesús Vioque2,4, Carmen Santamariña5, Carmen Pedraz-Pingarrón6, Soledad Abad7, María Ederra8, Dolores Salas-Trejo9, Carmen Vidal10, Carmen Sánchez-Contador11, Nuria Aragonés1,2,3, Beatriz Pérez-Gómez1,2,3 and Marina Pollán1,2,3 on behalf of DDM-Spain research group†

1Cancer Epidemiology Unit, National Center for Epidemiology, Instituto de Salud Carlos III, Avenida Monforte de Lemos, 5, 28029, Madrid, Spain
2Consortium for Biomedical Research in Epidemiology & Public Health (CIBERESP), Instituto de Salud Carlos III, Avenida Monforte de Lemos, 5, 28029, Madrid, Spain
3Cancer Epidemiology Research Group, Oncology and Hematology Area, IIS Puerta de Hierro (IDIPHIM), Calle Manuel de Falla, 1, 28222 Majadahonda, Madrid, Spain
4Universidad Miguel Hernandez, Crta. Nacional 332, s/n, 03550, Sant Joan D’Alacant, Alicante, Spain
5Galician Breast Cancer Screening Program, Galician Regional Health Authority, C/Gregorio Hernández, 2, 4, 15011, A Coruña, Spain
6Castile-León Breast Cancer Screening Program, General Directorate of Public Health, Avenida Sierra de Atapuerca, S/N, 09071, Burgos, Spain
7Aragón Breast Cancer Screening Program, Aragon Health Service, C/Ronda de Liberación, 1, 44002, Teruel, Zaragoza, Spain
8Navarre Breast Cancer Screening Program, Public Health Institute, C/ Leire, 15, 31003, Pamplona, Spain
9Valencian Breast Cancer Screening Program, General Directorate of Public Health, C/ Micer Mascó, 31, 46010, Valencia, Spain
10Cancer Prevention and Control Unit, Catalanion Institute of Oncology (ICO), Avenida Gran Vía, S/N, km 2.7, 08907, L'Hospitalet de Llobregat, Barcelona, Spain
11Balearic Islands Breast Cancer Screening Program, Regional Authority for Health & Consumer Affairs, C/ Cecilio Metelo, 18, 07012, Palma de Mallorca, Islas Baleares, Spain

(Submitted 14 December 2015 – Final revision received 25 April 2016 – Accepted 31 May 2016 – First published online 4 July 2016)

Abstract

The objective of the present study was to assess the reproducibility of data-driven dietary patterns in different samples extracted from similar populations. Dietary patterns were extracted by applying principal component analyses to the dietary information collected from a sample of 3550 women recruited from seven screening centres belonging to the Spanish breast cancer (BC) screening network (Determinants of Mammographic Density in Spain (DDM-Spain) study). The resulting patterns were compared with three dietary patterns obtained from a previous Spanish case–control study on female BC (Epidemiological study of the Spanish group for breast cancer research (GEICAM: grupo Español de investigación en cáncer de mama)) using the dietary intake data of 973 healthy participants. The level of agreement between patterns was determined using both the congruence coefficient (CC) between the pattern loadings (considering patterns with a CC ≥ 0.85 as fairly similar) and the linear correlation between patterns scores (considering as fairly similar those patterns with a statistically significant correlation). The conclusions reached with both methods were compared. This is the first study exploring the reproducibility of data-driven patterns from two studies and the first using the CC to determine pattern similarity. We were able to reproduce the EpiGEICAM Western pattern in the DDM-Spain sample (CC = 0.90). However, the reproducibility of the Prudent (CC = 0.76) and Mediterranean (CC = 0.77) patterns was not as good. The linear correlation between pattern scores was statistically significant in all cases, highlighting its arbitrariness for determining pattern similarity. We conclude that the reproducibility of widely prevalent dietary patterns is better than the reproducibility of more population-specific patterns. More methodological studies are needed to establish an objective measurement and threshold to determine pattern similarity.

Key words: Dietary patterns: Reproducibility: Congruence coefficients: Principal component analyses: Component loadings: Component scores

Abbreviations: BC, breast cancer; CC, congruence coefficient; DDM-Spain, Determinants of Mammographic Density in Spain; EpiGEICAM, Epidemiological study of the Spanish group for breast cancer research (GEICAM: grupo Español de investigación en cáncer de mama); PCA, principal component analysis.

* Corresponding author: Dr A. Castelló, fax +34 91 387 7815, email acastello@isciii.es
† Membership of the DDM-Spain research group is provided in the Acknowledgements section.
Diet is a key modifiable risk factor, but the exploration of its role in disease occurrence is complicated because of methodological issues related to the dietary assessment method used, food and nutrient interactions, and differences in food consumption across populations. Traditionally, nutritionists and researchers have explored the effect of individual dietary factors in disease occurrence. However, some authors advocate the use of dietary patterns instead of individual foods and nutrients, arguing that they may better capture variability in the population's diet, while allowing the evaluation of interactions between dietary factors.

These patterns can be identified with data-driven methods such as principal component analysis (PCA), factor analysis (FA) and cluster analysis or can be represented by investigator-driven patterns known as dietary quality indices. Investigator-driven patterns assign a set of scores based on individuals' fulfillment of a set of fixed recommendations. Therefore, they are widely applicable, facilitating the exploration of the reproducibility of their association with different diseases in independent populations. However, they present the disadvantage of being very disease dependent, given that they are mainly based on existing evidence of the association between diet and disease. On the other hand, data-driven dietary patterns are more representative of the diet of the specific population from which they have been extracted and independent of the diseases, but many authors argue that the patterns obtained are very population-dependent, and therefore difficult to reproduce in other settings.

The reproducibility of data-driven dietary patterns has been assessed previously by various authors using dietary information obtained with common assessment tools at different moments of time within the same sample. However, no previous studies have explored the reproducibility of data-driven dietary patterns extracted from different samples.

The objective of this study was to assess the reproducibility of data-driven dietary patterns in different samples extracted from similar populations. We compared the results from a previous cross-sectional, multicentre study carried out in seven screening centres belonging to the Spanish BC screening network and located throughout the Spanish peninsula. In Spain, all women aged 50–69 years (45–69 years in some regions), regardless of nationality or legal status, are invited to be screened under these government-sponsored programmes every 2 years. Women were randomly selected among all screening attendees and invited to participate on a daily basis until the minimum sample size of 500 for each centre was reached. A total of 3550 women were recruited between 2007 and 2008, with an average participation rate of 74.5% (range 64.7–84.0% across centres). Women were interviewed at the screening centres by trained interviewers who collected demographic, anthropometric, physical activity, gynaecologic, obstetric and occupational data, as well as family and personal history (including weight and height at age 18 years). Information on smoking included current status and months since quitting for ex-smokers. Current smokers were defined as women who smoked at the time of mammography or had quit <6 months before. Dietary intake during the preceding year was collected using a validated 117-item FFQ. Postmenopausal status was defined as self-reported absence of menstruation in the previous 12 months. Interviewers measured weight, height and waist and hip circumferences twice using the same protocol and identical balance scales, stadiometers and measuring tapes. A third measurement was taken when the first two were not equal.

The DDM-Spain study was conducted according to the guidelines laid down in the Declaration of Helsinki, and all procedures involving human subjects were approved by the bioethics and animal welfare committee at the Carlos III Institute of Health. All participants signed a consent form, including permission to publish results from the current research.

### Dietary patterns

The FFQ used in both studies were designed to assess the whole diet, had similar structures and were based on a validated FFQ. However, the FFQ of the DDM-Spain study included some additional food items that were not contained in the FFQ of the EpiGEICAM study: the FFQ used in the EpiGEICAM study contained ninety-nine items from which eighty-six were used to create the food groups (after excluding the non-energetic and alcoholic beverages), whereas the FFQ from DDM-Spain included 117 items (the same ninety-nine from DDM-Spain plus eighteen additional foods) from which ninety-nine were used to create the food groups (after excluding non-energetic and alcoholic beverages). In both cases, the dietary information collected was grouped into the exact same twenty-six food groups that are summarised in Table 1, where the items only included in the DDM-Spain study are represented in italics.

The EpiGEICAM study identified three dietary patterns over twenty-six food groups: a Western pattern characterised by elevated intakes of high-fat dairy products, processed meat, refined grains, sweets, energetic drinks and other convenience foods and sauces and by low intakes of low-fat dairy products and whole grains; a Prudent pattern defined by high intakes of low-fat dairy products, vegetables, fruits, whole grains and juices; and a Mediterranean pattern represented by a high intake of fish, vegetables, legumes, boiled potatoes, fruits, etc.
groups from the DDM-Spain sample. By applying the same PCA analysis to the same twenty-six food groups with pattern loadings, which indicate the importance of individual food groups in each pattern. Pattern weights and pattern loadings give similar information, except that they are measured on different scales (weights are standardised into scores from −1 to 1, and a value in the range 0.85–0.94 corresponds to fair similarity, whereas a value ≥0.95 implies that the two compared components/factors can be considered equivalent.

To evaluate the level of agreement between the food composition of patterns extracted in the DDM-Spain study and those reported in the EpiGEICAM study, we calculated congruence coefficients (CC) between the pattern loadings from both studies. CC represents the correlation between pattern loadings based on their deviations from 0 (instead of being based on the deviations from the mean of the factor loadings as the Pearson’s correlation is) and it is the preferred measure for component/factor similarity extracted with PCA/FA. CC ranges from −1 to 1, and a value in the range [0.85–0.94] corresponds to fair similarity, whereas a value ≥0.95 implies that the two compared components/factors can be considered equivalent.

For comparison purposes, we considered that food groups with pattern loadings ≥0.31 were the main contributors to a dietary pattern.

To evaluate the level of agreement between the food composition of patterns extracted in the DDM-Spain study and those reported in the EpiGEICAM study, we calculated congruence coefficients (CCs) between the pattern loadings from both studies. CC represents the correlation between pattern loadings based on their deviations from 0 (instead of being based on the deviations from the mean of the factor loadings as the Pearson’s correlation is) and it is the preferred measure for component/factor similarity extracted with PCA/FA. CC ranges from −1 to 1, and a value in the range [0.85–0.94] corresponds to fair similarity, whereas a value ≥0.95 implies that the two compared components/factors can be considered equivalent.
The CC between the pattern loadings of a given pattern from EpiGEICAM ($l_1$) and the pattern loadings of a given pattern from DDM-Spain ($l_2$) for each of the $j = 1, \ldots, 26$ food groups were calculated as follows:

$$CC = \frac{\sum_{j=1}^{26} l_1 j \times l_2 j}{\sqrt{\left(\sum_{j=1}^{26} l_1^2 j \right) \times \left(\sum_{j=1}^{26} l_2^2 j \right)}}.$$ 

In addition, to follow the same methodology commonly used in studies exploring the reproducibility of dietary patterns, Spearman’s correlation coefficients (Corr) between the EpiGEICAM and the DDM-Spain pattern scores were calculated. For that purpose, patterns scores (which reflect the level of compliance of each woman with each one of the dietary patterns) were calculated as the linear combination of consumption of food groups weighted by the pattern loadings from EpiGEICAM Western, Prudent and Mediterranean patterns and from the set of selected patterns resulting from applying PCA to the DDM-Spain data as follows:

$$P_{ki} = \sum_{j} (l_1 j \cdot C_{pj}),$$

where $P$ is the pattern score, $l$ the loading score, $C$ the centred food consumption, $k$ and $l$ the Western, Prudent and Mediterranean patterns from EpiGEICAM and Western, Prudent and Mediterranean patterns from DDM-Spain, $i = 1, \ldots, 3550$ women and $j = 1, \ldots, 26$ food groups.

CC is the preferred measure for component/factor similarity extracted with PCA/FA because its validity is supported by methodological research. In addition, a recent study has questioned the ability of using solely Pearson’s correlation (Corr) coefficient to assess pattern similarity. However, the majority of studies exploring the reproducibility of dietary patterns base their conclusions on the latter measure, considering any significant correlation as being indicative of pattern similarity regardless of its value. In this study, we provide the correlation coefficient for the sake of comparability with published data, but we will base our final conclusion regarding pattern reproducibility on the CC.

To take into account sampling variability in the estimation of pattern loadings using DDM-Spain data, and subsequently in the estimation of the agreement measurements between the patterns identified within the EpiGEICAM and the DDM-Spain studies, we performed a non-parametric bootstrap estimation with 5000 replications. Using sampling replacement, the bootstrap obtained 5000 replicates of the original DDM-Spain data set. PCA was then applied in each replication, and the three principal components that proved to be more similar to those reported in the EpiGEICAM were selected on the basis of the distance between the pattern loadings (more details are given in the online Supplementary Method 1). The 95% percentile CI for each parameter were represented by percentiles 2·5 and 97·5 of the 5000 bootstrap point estimates’ distribution.

Similar analyses were carried out by applying the PCA to food groups from the DDM-Spain study, which included the same exact eighty-six items considered in the EpiGEICAM analysis (online Supplementary Table S1 and Fig. S1).

Analyses were performed using STATA/MP 14.0.

Results
The anthropometric, reproductive and socio-demographic characteristics of the EpiGEICAM controls and DDM-Spain women are summarised in Table 2. The DDM-Spain study recruited a higher percentage of older and postmenopausal women (77 v. 47%), women with higher energy intake (on average 656 kJ/d (157 kcal/d) more in the DDM-Spain group), women with higher BMI and a higher percentage of women who practised physical activity with moderate-to-vigorous intensity (76 v. 63%). On the other hand, these women reported lower intake of alcohol, lower educational level (34% with primary school or less in DDM and 16% in EpiGEICAM), lower percentage of family history of BC (7 v. 20%), lower age at first delivery (43% of parous women in the DDM had their first child before 25 years of age, whereas this proportion was 26% in EpiGEICAM) and there was a lower percentage of nulliparous (9 v. 23%) women. The distribution of age at menarche and smoking appeared to be fairly similar in both studies.

Fig. 1–3 show the comparison between the original loadings from the EpiGEICAM study with their corresponding values in the DDM-Spain study. Western patterns from both studies were characterised by high intakes of high-fat dairy products, refined grains, energetic drinks and convenience food and sauces and low intakes of low-fat dairy products and whole grains. Correlations with the intake of red and/or processed meat and with sweets were also close to the 0·3 threshold. Moreover, the DDM-Spain Western pattern seemed to be negatively correlated with the consumption of white fish, a result that was not observed in EpiGEICAM. Despite these small differences, the elevated CC between patterns (CC = 0·90) indicates a fair similarity between the Western patterns extracted from the EpiGEICAM and the DDM-Spain data (Fig. 1).

We did not identify a pattern among women of the DDM-Spain study that was highly congruent with the EpiGEICAM Prudent pattern. The most similar pattern presented a high consumption of whole grains and juices but failed to correlate with low-fat dairy products, vegetables and fruits (Fig. 2). Something similar was observed with the Mediterranean pattern: several high correlations were observed with some vegetables, legumes, potatoes and nuts. However, the pattern from the DDM-Spain study did not include other typical factors of the Mediterranean diet, such as fish, olive oil and fruits (even if pattern loadings for these food groups were not low), whereas other foods more common in the Prudent diet, such as low-fat dairy products, or in the Western diet, such as sweets, and sugary and convenience foods, were included with high correlations. According to the CC (0·77), the EpiGEICAM and the DDM-Spain Mediterranean patterns cannot be considered similar (Fig. 3).

Finally, had we considered any significant correlation as being indicative of similarity, we would have concluded that all patterns extracted from the EpiGEICAM data were reproducible in the DDM-Spain study.

Discussion
To the best of our knowledge, this is the first study exploring the reproducibility of data-driven patterns in two different
samples extracted from similar populations. We were able to reproduce the Western pattern identified in women from the EpiGEICAM study among women attending BC screening programmes who participated in the DDM-Spain study. However, the reproducibility of the Prudent and Mediterranean patterns cannot be considered good.

The association between dietary patterns and BC has been explored in many studies in different settings. Most of these studies identified a Western/Unhealthy pattern, which shares the most important characteristics with the Western patterns identified in EpiGEICAM and DDM-Spain, such as high consumption of fatty dairy products, red/processed meat, refined grains, sweets and convenience foods. However, the Mediterranean and Prudent patterns have often been mixed under the names of Vegetable, Prudent, Healthy or Mediterranean diet. These patterns are characterised by a high consumption of vegetables and fruits that are an important part of the Mediterranean diet, but fail to include other items such as olive oil, nuts, legumes or fish, which are key foods to differentiate the so-called Prudent or Healthy patterns from the Mediterranean.

None of the above-mentioned studies have been able to identify both, a Prudent and a Mediterranean pattern in the same population, probably reflecting the difficulty in differentiating them in contexts where the Mediterranean diet is not very prevalent. On the other hand, the higher agreement in the definition of a Western pattern across studies is consistent with the greater reproducibility of this pattern observed in our study.

Table 2. Anthropometric, reproductive and socio-demographic characteristics of EpiGEICAM controls and Determinants of Mammographic Density in Spain (DDM-Spain) women
(Mean values and standard deviations; medians and interquartile ranges (IQR); numbers and percentages)

<table>
<thead>
<tr>
<th></th>
<th>EpiGEICAM controls (n 973)*</th>
<th>DDM-Spain (n 3550)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EpiGEICAM patterns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western pattern</td>
<td>0.00 3.77 16</td>
<td>0.00 2.31 16</td>
</tr>
<tr>
<td>Prudent pattern</td>
<td>0.00 3.34 13</td>
<td>0.34 (~2.21–1.92)†</td>
</tr>
<tr>
<td>Mediterranean pattern</td>
<td>0.00 2.70 8</td>
<td>0.00 1.50 7</td>
</tr>
<tr>
<td>Participants' characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy intake (kJ/d)</td>
<td>7937 2627</td>
<td>8593-93 2012-88</td>
</tr>
<tr>
<td>Energy intake (kcal/d)</td>
<td>1897 628</td>
<td>2054-15 481-09</td>
</tr>
<tr>
<td>Alcohol intake (g/d)</td>
<td>Median 2</td>
<td>0.85</td>
</tr>
<tr>
<td>IQR</td>
<td>0.04–7.10</td>
<td>0.0–5.68</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>25.36 4.28</td>
<td>28.03 5.99</td>
</tr>
<tr>
<td>Age (years)</td>
<td>50.63 9.47</td>
<td>56.20 5.46</td>
</tr>
<tr>
<td>Age at menarche (years)</td>
<td>12.44 1.52</td>
<td>13 12–14</td>
</tr>
</tbody>
</table>

Physical activity in the last year

<table>
<thead>
<tr>
<th></th>
<th>EpiGEICAM</th>
<th>DDM-Spain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>287 30</td>
<td>842 24</td>
</tr>
<tr>
<td>Moderate</td>
<td>368 38</td>
<td>1842 52</td>
</tr>
<tr>
<td>Vigorous</td>
<td>246 25</td>
<td>866 24</td>
</tr>
<tr>
<td>Unknown</td>
<td>72 7</td>
<td>–</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never or former &lt;6 months</td>
<td>645 67</td>
<td>2180 61</td>
</tr>
<tr>
<td>Smoker or former smoker &lt;6 months</td>
<td>325 33</td>
<td>1370 39</td>
</tr>
<tr>
<td>Unknown</td>
<td>3 0</td>
<td>–</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school or less</td>
<td>158 16</td>
<td>1204 34</td>
</tr>
<tr>
<td>Secondary school</td>
<td>489 50</td>
<td>1978 56</td>
</tr>
<tr>
<td>University</td>
<td>318 33</td>
<td>363 10</td>
</tr>
<tr>
<td>Unknown</td>
<td>8 1</td>
<td>5 0</td>
</tr>
<tr>
<td>Family history of BC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>782 80</td>
<td>3291 93</td>
</tr>
<tr>
<td>Yes</td>
<td>191 20</td>
<td>259 7</td>
</tr>
<tr>
<td>Age at first delivery (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>45 5</td>
<td>302 9</td>
</tr>
<tr>
<td>20–24</td>
<td>208 21</td>
<td>1194 34</td>
</tr>
<tr>
<td>25–29</td>
<td>266 27</td>
<td>1271 36</td>
</tr>
<tr>
<td>&gt;29</td>
<td>148 15</td>
<td>465 13</td>
</tr>
<tr>
<td>Nulliparous</td>
<td>220 23</td>
<td>316 9</td>
</tr>
<tr>
<td>Unknown</td>
<td>86 9</td>
<td>2 0</td>
</tr>
<tr>
<td>Menopausal status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Premenopausal</td>
<td>513 53</td>
<td>816 23</td>
</tr>
<tr>
<td>Postmenopausal</td>
<td>460 47</td>
<td>2734 77</td>
</tr>
</tbody>
</table>

v.e., Total variability in food group intakes explained by the pattern; BC, breast cancer.

* Descriptive data extracted from the scientific article of Castello et al. (24).
† As distribution of the prudent score was skewed, the median and IQR were used to describe this score.
Fig. 1. Pattern loadings of the Western dietary pattern extracted from the EpiGEICAM study (left) and pattern loadings and 95% percentile CI of the Western pattern extracted from Determinants of Mammographic Density in Spain (DDM-Spain) data (right). * Congruence coefficient (CC) and 95% percentile CI between EpiGEICAM and DDM-Spain pattern loadings. † Correlation coefficient (Corr) and 95% percentile CI between EpiGEICAM and DDM-Spain pattern scores. All correlations were significant at a 95% confidence level.

Fig. 2. Pattern loadings of the Prudent dietary pattern extracted from the EpiGEICAM study (left) and pattern loadings and 95% percentile CI of the Prudent pattern extracted from Determinants of Mammographic Density in Spain (DDM-Spain) data (right). * Congruence coefficient (CC) and 95% percentile CI between EpiGEICAM and DDM-Spain pattern loadings. † Correlation coefficient (Corr) and 95% percentile CI between EpiGEICAM and DDM-Spain pattern scores. All correlations were significant at a 95% confidence level.
As noted earlier in this study, PCA reduces a set of inter-correlated variables to a group of principal components (dietary patterns in this case) so that the maximum correlation between the variables within components and the minimum correlation among components are obtained. Therefore, the greater the variability in diet, the easier it will be to find clearly differentiated independent patterns. In our study, although EpiGEICAM included women from fourteen Spanish provinces (four of them on the Mediterranean coast), DDM-Spain participants were recruited from screening centres located in seven provinces (three of them located on the Mediterranean coast). Therefore, the greater geographical distribution in the EpiGEICAM study may imply a greater representativeness of all diets across the Spanish territory. In addition, distribution of age among DDM-Spain women was more homogeneous (range = 45–69) than that observed in the EpiGEICAM participants (range = 22–71). As García-Arenzana et al. previously described, older women tend to have healthier dietary habits than younger women, which may have produced a more heterogeneous distribution of dietary habits in the EpiGEICAM study. This heterogeneity might have facilitated the identification of more specific patterns, not only limited to the discrimination of two antagonistic patterns (Western vs. Healthy/Prudent/Mediterranean) but also allowing the clear differentiation of patterns with subtle differences, such as the Prudent and Mediterranean patterns.

Regarding the pre-established thresholds for the CC that define the similarity of dietary patterns in both studies, we based our decision on three published pieces of research that evaluated concordance coefficients in light of the subjective opinion of several experienced researchers judging the equivalence between different components. Haven and Nesselroade argue that values over 0.80 are enough to assume fair similarity between components, whereas Lorenzo-Seva & Berge maintain a more conservative approach setting the cut-off point for fair similarity at 0.85 and preventing a CC below this value from being interpreted as indicative of similarity. All three articles agree on the difficulty in setting up a cut-off point under which patterns should be considered clearly different. Despite the fact that the CC is considered a good measure of agreement between components or factors extracted with PCA or FA, the existing bibliography evaluating the reproducibility of data-driven dietary patterns does not use this measure and bases its conclusions only on the correlations between pattern scores, considering any significant correlation as being indicative of similarity regardless of its magnitude.

![Diagram showing pattern loadings of the Mediterranean dietary pattern extracted from the EpiGEICAM study (left) and pattern loadings and 95% percentile CI of the Mediterranean pattern extracted from Determinants of Mammographic Density in Spain (DDM-Spain) data (right).](https://doi.org/10.1017/S000711451600252X)
but also are similarly associated with BC risk. 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. In the present study, taking into account only the methodological studies published regarding the threshold of the CC for pattern similarity, we followed the most conservative approach and considered dietary patterns to be fairly similar if CC values were ≥0.85.

A major limitation of the use of dietary patterns 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 as follows: 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 assignment of a label to each of these patterns. However, we have demonstrated that this limitation can be overcome by a detailed analysis when comprehensive information on food grouping and loadings is provided by Castello et al. On the other hand, both FFQ from EpiGEICAM and DDM-Spain collected information on ninety-nine identical foods, except for the fact that DDM-Spain included eighteen additional foods that were not included in EpiGEICAM. In addition, the same group of researchers took principal responsibility for the analysis of the data; therefore, food grouping and labelling were very similar in both studies.

Finally, we summarise the main strengths of the present study. As previously mentioned, various studies have assessed the reproducibility of investigator-driven patterns. The reproducibility of data-driven dietary patterns extracted from the same sample using the dietary information obtained with different assessment tools or in different time points, the thresholds chosen to determine the contribution of food groups to the identified dietary patterns and the assignment of a label to each of these patterns. However, we have demonstrated that this limitation can be overcome by a detailed analysis when comprehensive information on food grouping and loadings is provided by Castello et al. On the other hand, both FFQ from EpiGEICAM and DDM-Spain collected information on ninety-nine identical foods, except for the fact that DDM-Spain included eighteen additional foods that were not included in EpiGEICAM. In addition, the same group of researchers took principal responsibility for the analysis of the data; therefore, food grouping and labelling were very similar in both studies.

Conclusions

The reproducibility of widely prevalent dietary patterns such as the Western pattern is better than the reproducibility of patterns more specific to certain populations, such as the Mediterranean. More methodological studies exploring the reproducibility of dietary patterns are needed to establish a more objective threshold for the CC between pattern loadings and their equivalent Corr between pattern scores that define pattern similarity.

Acknowledgements

The authors thank the DDM-Spain study participants for their contribution to breast cancer research and all collaborator researchers: Pilar Moreo, María Pilar Moreno, María Soledad Abad, Francisca Collado, Francisco Casanova, Jose Antonio Vázquez, Nieves Asuncion, Milagros García, Manuela Alcaraz, María Soledad Laso, Josefa Miranda and Francisco Ruiz Perales.

This study was supported by Carlos III Institute of Health FIS (Spanish Public Health Research Fund: PI060386 FIS; PS09/00790 and PI15/10029 research grants), the Spanish Ministry of Health (EC11-275), the Spanish Ministry of Economy and Competitiveness (IJC-2014-20900), the Spanish Federation of Breast Cancer Patients (FECMA: EPI 1169-10) and the Association of Women with Breast Cancer from Elche (AMAC-MEC: EPI 1394/15). None of the funders had any role in the design, analysis or writing of this article.

V. L., N. A., B. P.-G. and M. P. designed the study; A. C., J. V., C. S., C. P.-P., S. A., M. E., D. S.-T., C. V. and C. S.-C. collected the data and/or prepared the database. A. C. performed statistical analysis and wrote the initial version of the manuscript that M. P. revised and corrected in its different versions. All the authors have read and approved the final version of the manuscript.

The authors declare that there are no conflicts of interest.

Supplementary material

For supplementary material/s referred to in this article, please visit http://dx.doi.org/10.1017/S000711451600252X


