This is the published version of a paper published in *Journal of Nutrition*.

Citation for the original published paper (version of record):

Generalizability of a Diabetes-Associated Country-Specific Exploratory Dietary Pattern Is Feasible Across European Populations
*Journal of Nutrition*, 149(6): 1047-1055
https://doi.org/10.1093/jn/nxz031

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-160601
Generalizability of a Diabetes-Associated Country-Specific Exploratory Dietary Pattern Is Feasible Across European Populations

Franziska Jannasch,1,2,3 Janine Kröger,1,2 Claudia Agnoli,4 Aurelio Barriacarte,5,6,7 Heiner Boeing,8 Valerie Cayssials,9 Sandra Colorado-Yohar,5,10,11 Christina C Dahm,12 Courtney Dow,13 Guy Fagherazzi,13 Paul W Franks,14,15 Heinz Freisling,16 Marc J Gunter,16 Nicola D Kerrison,17 Timothy J Key,18 Kay-Tee Khaw,19 Tilman Kühn,15 Cecile Kyro,22 Francesca Romana Mancini,13 Olaf Mokoroa,22 Peter Nilsson,24 Kim Overvad,1,2,21 Domenico Palli,24 Salvatore Panico,25 Jose Ramón Quirós García,26 Olov Rolandsson,13 Carlotta Sacerdote,27,28 Mariá-José Sánchez,29 Mohammad Sedij Sahrai,18 Ruth Schübel,20 Ivonne Slujs,30 Annemieke MW Spijkerman,31 Anne Tjonneland,21 Tammy YN Tong,18 Rosario Tumino,32,33 Elio Riboli,34 Claudia Langenberg,17 Stephen J Sharp,17 Nita G Forouhi,17 Matthias B Schulze,1,2,3 and Nicholas J Wareham17

1Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Germany; 2German Center for Diabetes Research (DZD), München-Neuherberg, Germany; 3NutriAct—Competence Cluster Nutrition Research Berlin-Potsdam, Nuthetal, Germany; 4Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; 5CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain; 6Navarra Public Health Institute, Pamplona, Spain; 7IdaNA, Navarra Institute for Health Research, Pamplona, Spain; 8Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany; 9Unit of Nutrition and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain; 10Department of Epidemiology, Murcia Health Council, IMIB-Arraxaca, Spain; 11Research Group on Demography and Health, National Faculty of Public Health, University of Antioquia, Medellin, Colombia; 12Section for Epidemiology, Department of Public Health, Aarhus University, Denmark; 13Inserm, Centre for Research in Epidemiology and Population Health [CERS], Villejuif, France; University Paris-Sud, Villejuif, France; 14Department of Clinical Sciences, Lund University, Malmo, Sweden; 15Department of Public Health and Clinical Medicine, Umeå University, Sweden; 16Section of Nutrition and Metabolism, International Agency for Research on Cancer (IARC-WHO), Lyon, France; 17MRC Epidemiology Unit, University of Cambridge, United Kingdom; 18Cancer Epidemiology Unit, University of Oxford, Oxford, United Kingdom; 19Department of Public Health and Primary Care, University of Cambridge, Addenbrooke's Hospital, Cambridge, United Kingdom; 20German Cancer Research Center [DKFZ], Heidelberg, Germany; 21Danish Cancer Society Research Center, Copenhagen, Denmark; 22Public Health Division of Gipuzkoa, Biodonostia Research Institute, San Sebastian, Spain; 23Department of Cardiology, Center for Cardiovascular Research, Aalborg Hospital, Aarhus University Hospital, Aalborg, Denmark; 24Cancer Risk Factors and Life-Style Epidemiology Unit, Institute for Cancer Research, Prevention and Clinical Network—ISPRO, Florence, Italy; 25Department of Medicina Clinica e Chirurgia, Federico II University, Naples, Italy; 26Public Health Directorate, Asturias, Spain; 27Unit of Cancer Epidemiology, Città della Salute e della Scienza Hospital-University of Turin and Center for Cancer Prevention (CPO), Torino, Italy; 28Human Genetics Foundation (HuGeF), Turin, Italy; 29Andalusian School of Public Health, Granada, Spain; 30Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands; 31National Institute for Public Health and the Environment [RIVM], Bilthoven, Netherlands; 32Cancer Registry and Histopathology Unit, ‘Civile – M.P. Arezzo’ Hospital, Ragusa, Italy; 33Associazione Iblea per la Ricerca Epidemiologica—Onlus, Ragusa, Italy; 34Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, United Kingdom and 35Institute of Nutritional Sciences, University of Potsdam, Nuthetal, Germany

ABSTRACT

Background: Population-specificity of exploratory dietary patterns limits their generalizability in investigations with type 2 diabetes incidence.

Objective: The aim of this study was to derive country-specific exploratory dietary patterns, investigate their association with type 2 diabetes incidence, and replicate diabetes-associated dietary patterns in other countries.

Methods: Dietary intake data were used, assessed by country-specific questionnaires at baseline of 11,183 incident diabetes cases and 14,694 subcohort members (mean age 52.9 y) from 8 countries, nested within the European Prospective Investigation into Cancer and Nutrition study (mean follow-up time 6.9 y). Exploratory dietary patterns were derived by principal component analysis. HRs for incident type 2 diabetes were calculated by Prentice-weighted Cox proportional hazard regression models. Diabetes-associated dietary patterns were simplified or replicated to be applicable in other countries. A meta-analysis across all countries evaluated the generalizability of the diabetes-association.

Results: Two dietary patterns per country/UK-center, of which overall 3 dietary patterns were diabetes-associated, were identified. A risk-lowering French dietary pattern was not confirmed across other countries: pooled HRFrance per 1 SD: 1.00; 95% CI: 0.90, 1.10. Risk-increasing dietary patterns, derived in Spain and UK-Norfolk, were confirmed, but only the latter statistically significantly: HRSpain: 1.09; 95% CI: 0.97, 1.22 and HRUK-Norfolk: 1.12; 95% CI: 1.04, 1.20. Respectively,

Copyright © American Society for Nutrition 2019. All rights reserved. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com


First published online April 24, 2019; doi: https://doi.org/10.1093/jn/nxz031
this dietary pattern was characterized by relatively high intakes of potatoes, processed meat, vegetable oils, sugar, cake and cookies, and tea.

**Conclusions:** Only few country/center-specific dietary patterns (3 of 18) were statistically significantly associated with diabetes incidence in this multicountry European study population. One pattern, whose association with diabetes was confirmed across other countries, showed overlaps in the food groups potatoes and processed meat with identified diabetes-associated dietary patterns from other studies. The study demonstrates that replication of associations of exploratory patterns with health outcomes is feasible and a necessary step to overcome population-specificity in associations from such analyses. J Nutr 2019;149:1047–1055.

**Keywords:** dietary patterns, principal component analysis, diet-disease association, type 2 diabetes mellitus, replication, meta-analysis

---

**Introduction**

Numerous studies have determined the association between diet and incident type 2 diabetes and have summarized evidence of the properties of specific single food groups that promote or reduce the risk. For instance, processed meat intake and sugar-sweetened beverages have been associated with greater diabetes risk and wholegrain intake with lower intake and sugar-sweetened beverages have been associated with type 2 diabetes (1–4). However, single food group investigations do not consider the complexity of human diet and the potential interplay of nutrients (5). To address these limitations, dietary patterns can serve as an alternative approach to examine which combinations of foods potentially contribute to development of noncommunicable diseases such as type 2 diabetes. If no previous knowledge about overall dietary patterns in a particular population is available, exploratory methods such as principal component analysis (PCA) and cluster analysis, or mixed approaches such as reduced rank regression, give a first insight. However, these approaches result by definition in population-specific dietary patterns with likely limited comparability (6). Indeed, when we recently summarized prospective studies on exploratory patterns and diabetes risk (7), heterogeneity in the pattern structure limited attempts to meta-analyze evidence. Given that single-study findings are unlikely to inform dietary recommendations, the lack of replication of dietary pattern-disease associations remains a major limitation. This might be particularly challenging for European populations with a high degree of heterogeneity of diets (8, 9).

Although methodological approaches to replicate diet-disease associations for exploratory patterns exist and have been used for patterns derived by reduced rank regression (7, 10), we are not aware of studies that have replicated population-specific patterns from factor analysis or PCA. Because methodological differences in dietary assessments might impact the ability to replicate dietary pattern-disease associations (11), multicenter studies across different populations with standardized dietary assessment and harmonized data to minimize heterogeneity would be very useful in this context. The EPIC (European Prospective Investigation into Cancer and Nutrition)-InterAct study provides a promising setting, as for the collection of dietary information, special attempts were made to harmonize the data across the countries (12). Thus, the study sets a perfect starting point to derive exploratory dietary patterns and to the potential for replication. Hence, the aim of the current investigation was the identification of country-specific dietary patterns with PCA and their association with type 2 diabetes risk. Subsequently, the potential for...
replication of identified diabetes-associated dietary patterns across all EPIC-InterAct countries was evaluated to overcome the population-specificity of exploratory dietary patterns.

**Methods**

**Study population and design**

The EPIC-InterAct study is a case-cohort study, nested within the prospective EPIC study (12). Between 1992 and 2000, more than a half million (n = 519,978) participants with an age range of 35–70 y were enrolled for the EPIC study. Participants were recruited from the general population despite some exceptions (13–15). Written informed consent was provided by each participant and the study was approved by local ethics committees and the Internal Review Board of the International Agency for Research on Cancer (14). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

The EPIC-InterAct case-cohort was established based on the incident cases of type 2 diabetes in the full (EPIC) cohort, which occurred between 1991 and 2007 in 8 of 10 participating countries (except for Norway and Greece) and a randomly drawn subcohort (12). From 455,680 eligible participants, those without stored blood samples (n = 109,625) and reported diabetes status (n = 5821) were excluded, resulting in 340,234 participants. A subcohort of 16,835 participants was randomly selected, stratified by center. After exclusion of post-censoring diabetes cases and individuals with unknown status (n = 681), the subcohort comprised 16,154 participants (Supplemental Figure 1).

Across the EPIC centers, different sources of evidence were used for ascertainment of incident cases of type 2 diabetes, including self-report, linkage to primary-care or secondary-care registers, medication use (drug registers), hospital admissions, or mortality data. Any of these sources of information was acceptable, where the information was obtained after the date of baseline examination. In Denmark and Sweden, diabetes cases were identified via local and national diabetes and pharmaceutical registers and therefore were considered to be verified. In other centers, if information on diabetes status was ascertained from only 1 source of evidence, further evidence to specify the definition for these cases was sought by including reviews of medical records. Censoring of the follow-up was either done at the date of diagnosis, 31 December 2007 or the date of death, depending on the first occurrence (12).

For the group of ascertained type 2 diabetes cases (n = 17,928), exclusion criteria similar to the subcohort were applied and further self-reports of diabetes in Denmark and nondiabetic participants were excluded (n = 5523), resulting in 12,403 verified type 2 diabetes cases (12).

In the Swedish study center Umeå, information on certain food groups (vegetables, dairy, other fruits, meat, offal, eggs, and vegetable oils) was not available, which was needed for the principal component analysis. Therefore, data from this study center were excluded (n = 1845). Further exclusions on missing food groups, anthropometry, and lifestyle factors such as smoking or physical activity were applied (n = 776), resulting in a subcohort of 14,694 participants and 11,183 verified diabetes cases. Because of the random selection of the subcohort members, 719 type 2 diabetes cases overlapped with the subcohort (Supplemental Figure 1). Excluded participants (n = 2621) were slightly younger, more likely to be men, with a higher BMI and waist circumference, less physically active, with a lower proportion of highly educated participants, a higher proportion of family history of diabetes and higher glycated hemoglobin (HbA1c), but a lower total energy intake (Supplemental Table 1).

**Dietary assessment**

Country-specific dietary questionnaires, developed and locally validated in a series of studies (16, 17), were used to assess the usual food intake over the previous 12 mo. Although the reproducibility was generally good, the validity ranged from modest to good (17).

Food items were aggregated in a common food classification system (Supplemental Table 2), which was based on the EPIC-Soft classification system (18), but provided a higher level of detail (more subgroups) to be used for dietary pattern analyses. For the purpose of an EPIC-InterAct wide application of principal component analysis, the food groups “saucers,” “soups,” and “miscellaneous” were not considered as they were not available in all EPIC centers. Total energy intake and intake of specific nutrients were derived from the “EPIC Nutrient Database” (19).

**Assessment of covariates**

Standardized self-report questionnaires on sociodemographic and lifestyle information such as age, education, smoking history, physical activity, and history of previous illnesses were used in all EPIC countries (14). Standardized protocols were used to measure anthropometric data including height, weight, waist and hip circumference. In France and Oxford, self-reports of all 4 anthropometric variables were obtained and adjusted by measured values from a subset of participants (14).

**Statistical methods**

Baseline characteristics of the participants in the subcohort were stratified by country and presented as mean ± SD, if normally distributed, median and interquartile range, if not normally distributed, and for categorical variables as relative percentages. The UK cohorts from Norfolk (population-based) and Oxford (high proportion of vegetarians, vegans, and health-conscious individuals) were considered separately. Country-specific dietary patterns in the subcohort were derived with PCA, considering the eigenvalue >1 criterion, scree plot, and interpretability to decide for the final number of principal components (20). Given the assumption that each food group has a variance of 1 in the PCA, retaining only those principal components with an eigenvalue (variance explained by 1 principal component) >1 reduces the dimensionality. In a scree plot these eigenvalues were plotted against all principal components (number equals the number of food groups) to visualize a possible “screw” in the top-to-ground slope. Principal components above the screw could be identified as explaining the majority of variance. Subsequently, the interpretability criterion (defined as ≥3 food groups with absolute factor loadings ≥0.4) was applied on these retained principal components to identify only those that reflect the complexity of a dietary pattern. In PCA, factor loadings can be interpreted as correlations between the food groups and resulting principal components (20). Then, dietary pattern scores were calculated for subcohort participants in each country as the sum of country-specific weights (βi, i = 1,...,m) equivalent to factor loadings, multiplied by food groups (Xi, i = 1,...,m) (standardized to respective country-specific subcohort distributions) (Equation 1) (Step 1 in Figure 1).

\[
\text{Dietary pattern score} = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_m X_m
\] (1)

Subsequently, dietary pattern scores were calculated for the cases external to the subcohort in the respective country with use of standardized food groups according to Equation 1. Cox proportional hazard regression models, weighted according to Prentice (21) and stratified by center and integers of age (y), were performed to investigate the association between the dietary patterns and the risk of type 2 diabetes within the whole case-cohort in each country (Step 2 in Figure 1). Two levels of adjustment were applied: Model 1 was adjusted for sex, physical activity [classified into “inactive,” “moderately inactive,” “moderately active,” and “active” according to the validated Cambridge Physical Activity Index (22)], educational level (none, primary, technical/professional, secondary, university), smoking status (never, former, current), and total energy intake (continuous). Model 2 was also adjusted for BMI (continuous) and waist circumference (continuous).

To facilitate the applicability of the identified diabetes-associated country-specific dietary patterns to other study populations, a simplification approach (6) was applied. For this purpose, simplified sum scores were constructed by summing only those unweighted food groups (Xi,
i = 1, ..., n), which had high factor loadings in the original country-specific dietary patterns (6) (Equation 2). The food groups were again standardized to the country-specific subcohort distribution.

\[
\text{Simplified sum score} = X_1 + X_2 + X_3 + \ldots + X_m \quad (n < m) \tag{2}
\]

Because no common cutoff has been defined as “high” factor loading, we applied 4 different cutoffs for absolute factor loadings ≥0.4, ≥0.3, ≥0.2, and ≥0.1 to calculate simplified sum scores. Depending on the chosen cutoff for the factor loading, the number of food groups (n) to sum up differed, but was always smaller than all 36 food groups (m). The resulting simplified sum scores were standardized according to the distribution of the country-specific subcohorts. The original dietary pattern score (Equation 1) was sufficiently reflected by the simplified sum score (Equation 2), if Spearman correlation coefficients of \( r \geq 0.90 \) were achieved. However, if \( r < 0.90 \) within the respective country, we did not consider the use of simplified sum scores, but calculated replicative scores in the other countries instead: sum of the product of all 36 standardized food groups (\( X_n \)) multiplied by standardized scoring coefficients (\( \beta_n \)). EPIC, European Prospective Investigation into Cancer and Nutrition.

![Table 1](https://academic.oup.com/jn/article-abstract/149/6/1047/5478921)

**FIGURE 1** Scheme of the statistical steps to investigate the association between dietary patterns and type 2 diabetes risk. Subcohort in each country = randomly drawn subcohort in each country including noncases and cases of type 2 diabetes; whole case-cohort in each country = randomly drawn subcohort in each country and type 2 diabetes cases external to the subcohort; whole EPIC-InterAct case-cohort = sum of all randomly drawn subcohorts (\( n = 14,694 \)) and external type 2 diabetes cases (\( n = 11,183 \)), with an overlap of \( n = 719 \) verified incident type 2 diabetes cases in the subcohort, across all included EPIC-InterAct countries. *Sum of a reduced number (\( m < n \)) of standardized, unweighted food groups characterized by high factor loadings in the original dietary pattern score. **Sum of all 36 standardized food groups (\( X_n \)) multiplied by standardized scoring coefficients (\( \beta_n \)).
Results

Characterization of the study population

Country-specific baseline characteristics of the EPIC-InterAct subcohort (n = 14,694) are shown in Table 1. The participants were middle-aged, with wide ranges for BMI and waist circumference, and with a rather low percentage of participants being physically active and a low percentage of participants having post-secondary education in most EPIC-InterAct countries.

Relative contributions of macronutrients to the total energy intake were comparable across the countries. Except for few food groups, consistently consumed in small amounts in all countries, for example, “other fruits” and nuts, food intake largely varied between countries. Some food groups appeared to represent a country-specific intake, for example, a 3-fold higher intake of legumes in Spain compared to other countries or “other cereals” (including flour, breakfast cereals, and salty biscuits, among others) in the 2 UK centers. By contrast, some countries were characterized by a very low intake of specific food groups, for example, intake of leafy vegetables in Denmark.

Country-specific dietary patterns and their association with type 2 diabetes risk

In each EPIC-InterAct country, 2 dietary patterns were identified with PCA according to the defined criteria (Supplemental Table 4). The structure of dietary patterns, thus the factor loadings of the 36 included food groups, showed substantial heterogeneity between countries. We subsequently evaluated these dietary patterns within the countries in which they were derived.

Of the 18 identified dietary patterns, 3 were significantly associated with diabetes risk in the most adjusted models (Table 2): in Norfolk, the dietary pattern 2 explained 7.5% and was characterized by high factor loadings of potatoes, processed meat, vegetable oils, sugar, cakes and cookies, and tea.

Exploratory dietary patterns and type 2 diabetes

1 Model 1 was adjusted for sex, physical activity, smoking, education, and total energy intake; Model 2 was further adjusted for BMI and waist circumference.

### Table 1: Country-specific baseline characteristics of the EPIC-InterAct subcohort (n = 14,694)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>France (n = 562)</th>
<th>Italy (n = 1927)</th>
<th>Spain (n = 3509)</th>
<th>UK-Norfolk (n = 900)</th>
<th>UK-Oxford (n = 316)</th>
<th>Netherlands (n = 1398)</th>
<th>Germany (n = 2044)</th>
<th>Sweden (n = 1919)</th>
<th>Denmark (n = 2119)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>56.3 ± 6.5</td>
<td>50.2 ± 7.9</td>
<td>49.1 ± 7.8</td>
<td>59.5 ± 9.4</td>
<td>50.2 ± 11.6</td>
<td>52.6 ± 10.8</td>
<td>50.3 ± 8.6</td>
<td>57.8 ± 7.7</td>
<td>56.6 ± 4.4</td>
</tr>
<tr>
<td>Men, %</td>
<td>0</td>
<td>32.3</td>
<td>38.4</td>
<td>42.7</td>
<td>27.9</td>
<td>16.7</td>
<td>41.5</td>
<td>40.3</td>
<td>53.7</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>23.1 ± 3.7</td>
<td>25.8 ± 4.1</td>
<td>28.2 ± 4.2</td>
<td>25.7 ± 3.7</td>
<td>24.3 ± 3.9</td>
<td>25.3 ± 3.9</td>
<td>25.8 ± 4.1</td>
<td>25.3 ± 3.9</td>
<td>26.0 ± 3.9</td>
</tr>
<tr>
<td>WC, cm</td>
<td>—</td>
<td>91.9 ± 9.7</td>
<td>99.2 ± 8.7</td>
<td>92.7 ± 9.6</td>
<td>89.0 ± 11.4</td>
<td>91.3 ± 11.2</td>
<td>94.4 ± 9.7</td>
<td>93.4 ± 10.3</td>
<td>95.9 ± 9.8</td>
</tr>
<tr>
<td>Men</td>
<td>9.6</td>
<td>14.8</td>
<td>11.4</td>
<td>14.8</td>
<td>15.2</td>
<td>40.6</td>
<td>20.3</td>
<td>16.4</td>
<td>35.4</td>
</tr>
<tr>
<td>Never smoking, %</td>
<td>64.6</td>
<td>45.0</td>
<td>54.9</td>
<td>46.9</td>
<td>58.9</td>
<td>40.2</td>
<td>47.1</td>
<td>39.1</td>
<td>33.7</td>
</tr>
<tr>
<td>Post-secondary education, %</td>
<td>39.9</td>
<td>14.5</td>
<td>11.3</td>
<td>12.3</td>
<td>42.4</td>
<td>21.7</td>
<td>34.9</td>
<td>22.4</td>
<td>20.4</td>
</tr>
<tr>
<td>Mean Age, y</td>
<td>562</td>
<td>1927</td>
<td>3509</td>
<td>900</td>
<td>316</td>
<td>1398</td>
<td>2044</td>
<td>1919</td>
<td>2119</td>
</tr>
<tr>
<td>Mean Education, %</td>
<td>96</td>
<td>14.8</td>
<td>11.4</td>
<td>14.8</td>
<td>15.2</td>
<td>40.6</td>
<td>20.3</td>
<td>16.4</td>
<td>35.4</td>
</tr>
<tr>
<td>Mean Never smoking, %</td>
<td>64.6</td>
<td>45.0</td>
<td>54.9</td>
<td>46.9</td>
<td>58.9</td>
<td>40.2</td>
<td>47.1</td>
<td>39.1</td>
<td>33.7</td>
</tr>
<tr>
<td>Mean Post-secondary education, %</td>
<td>39.9</td>
<td>14.5</td>
<td>11.3</td>
<td>12.3</td>
<td>42.4</td>
<td>21.7</td>
<td>34.9</td>
<td>22.4</td>
<td>20.4</td>
</tr>
</tbody>
</table>

1 Data are shown as means ± SDs or as relative percentages. EPIC, European Prospective Investigation into Cancer and Nutrition; WC, waist circumference.
was not associated with risk of diabetes in Model 1 (HR: 1.15; 95% CI: 0.92, 1.42); however, adjustment for anthropometric markers (Model 2) indicated an inverse association (HR: 0.64; 95% CI: 0.49, 0.85) (Table 2). The dietary pattern 1 in Spain was characterized by high contributions of potatoes, legumes, bread, red meat, processed meat, eggs, vegetable oils, wine and spirits, and explained 9.7% of the total variance in the food groups (Supplemental Figure 4). It was associated with an increased diabetes risk in Model 1 (HR: 1.32; 95% CI: 1.19, 1.46), whereas the HR was markedly attenuated, when also adjusted for anthropometric markers (HR: 1.14; 95% CI: 1.03, 1.27) (Table 2).

**Replication of selected country-specific dietary patterns across other countries**

Those dietary patterns significantly associated with type 2 diabetes in Model 2 (Table 2), were subsequently replicated in all other countries. As a first step, the Spearman correlation coefficients of the simplified sum scores (by step-wise lowering the cutoff for the factor loadings) with the original dietary patterns were moderate in France (r = 0.59–0.69) and Norfolk (r = 0.66–0.72) (Supplemental Table 5) and judged not to sufficiently reflect the original dietary patterns. Hence, replicative scores were calculated instead. In contrast, the simplified sum score in the Spanish subcohort showed good correlations with the original scores ranging from r = 0.90–0.93. As lowering the cutoff of absolute factor loadings from ≥0.3 to ≥0.2 did not result in a markedly improved correlation, but in a highly increased number of food groups, the simplified sum score consisting of those foods with loadings ≥0.3 (potatoes, legumes, bread, red meat, processed meat, eggs, vegetable oils, wine, spirits, and pasta and rice) was considered.

Subsequently, the simplified and replicative scores were calculated in all EPIC-InterAct countries and evaluated with regard to diabetes risk. The “Replicative Norfolk” score was associated with an increased diabetes risk in the meta-analysis (Figure 2) (pooled HR for 1 SD increment, model 2 adjustments: 1.12; 95% CI: 1.04, 1.20). Inconsistency between the countries was existent (I² = 50%), although lower than for the other 2 replicated scores. With the exception of France and Spain, the positive direction of an association with diabetes risk in all other countries could be replicated, although not always significantly. To investigate whether the pooled estimate in the meta-analysis was driven by few countries, we excluded each country in a sensitivity analysis. The exclusion of Spain resulted in a strengthened association (HR: 1.16; 95% CI: 1.10, 1.23), whereas the exclusion of Denmark attenuated the pooled estimate (HR: 1.09; 95% CI: 1.02, 1.18). Hence, the overall clear positive association remained.

For the “Replicative France” score, the meta-analysis resulted in a pooled HR: 1.00; 95% CI: 0.90, 1.10 (Supplemental Figure 5), thus not confirming the inverse association of the Component 2 observed initially in France. The inconsistency measure indicated high heterogeneity between countries (I² = 67%). A clear inverse association with type 2 diabetes risk was exclusively observed in France, whereas a clear risk increase of 44% was observed in UK-Norfolk.

For the “Simple Spain” score the pooled HR: 1.09; 95% CI: 0.97, 1.22 and inconsistency between the countries was high (I² = 74%) (Supplemental Figure 6). A higher diabetes risk was observed in UK-Norfolk, besides Spain. Although in the other countries the simplified score was not significantly associated with incident type 2 diabetes, a clear inverse association was observed in Italy. A sensitivity analysis was conducted to investigate whether the exclusion of wine would result in a stronger positive association with diabetes risk. However, the pooled estimate was not substantially changed (HR: 1.08; 95% CI: 0.98, 1.20).

We subsequently applied several sensitivity analyses to evaluate the robustness of our findings. Excluding participants with a history of cardiovascular diseases, for example, myocardial infarction, stroke, or angina (n = 1195), who developed type 2 diabetes within the first 2 y of follow-up and with baseline HbA1c ≥6.5% (n = 2492), did not change the overall results [pooled HR (Norfolk): 1.11; 95% CI: 1.02, 1.20, I² = 42%; pooled HR (France): 0.98; 95% CI: 0.88, 1.09, I² = 75%; pooled HR (Spain): 1.10; 95% CI: 0.98, 1.23, I² = 75%]. Exclusion of participants in the top and bottom 1% of the energy intake/energy requirement ratio (n = 551) did not alter the results [pooled HR (Norfolk): 1.10; 95% CI: 1.05, 1.15, I² = 14%; pooled HR (France): 1.00; 95% CI: 0.91, 1.09, I² = 69%; pooled HR (Spain): 1.08; 95% CI: 0.97, 1.21, I² = 74%]. Adjusting for family history of type 2 diabetes (n = 3102) in countries where information was ascertained, gave the same diabetes-associated dietary patterns as in the main analyses.

**Discussion**

In this European case-cohort study, 2 dietary patterns per country were identified by PCA with highly specific structures.
of contributing food groups. Only a small proportion (3 of 18 dietary patterns) was significantly associated with diabetes risk (1 in France, 1 in Spain, 1 in UK-Norfolk). Of these dietary patterns, the positive association of the UK-Norfolk pattern could potentially be replicated across other countries in the EPIC-InterAct study. In contrast, no meaningful association could be established for the pattern from France, given the heterogeneous findings in terms of direction of association and overall null association from the meta-analysis. For the Spanish pattern, the positive association with diabetes risk could not be replicated, although in 5 of 8 countries the HRs pointed towards a positive association, even if the significance level was not reached.

The “Replicative Norfolk” score was quite consistently associated with an increased diabetes risk across most countries, as confirmed by the sensitivity analyses, where the overall association was not mainly driven by a single country contribution. Food groups with high positive factor loadings were potatoes, processed meat, vegetable oils, sugar, cake and cookies, and tea. Among these, evidence for risk-increasing properties exists for processed meat (24) and partly for potatoes (including French fries) (25). In previous EPIC-InterAct analyses, cakes and cookies, as well as vegetable oils, were not associated with diabetes risk when adjusted for BMI (26). Olive oil intake, in particular, contributed highly to lower diabetes risk (27). However, vegetable oil intake in non-Mediterranean countries is potentially driven by contributions of oils other than olive oil (26). Overall, these data suggest that vegetable oils are unlikely to explain the positive association of the “Replicative Norfolk” score with diabetes risk. For tea consumption, a systematic review summarized an inverse association in Asian studies, but no association in European and American studies (28). However, a previous publication in EPIC-InterAct showed an inverse association between the consumption of ≥4 cups of tea/d and diabetes risk (29). Our food group “sugar” included confectionary, honey, jams, chocolate, syrups, and ice cream. This heterogeneous mix of food items is characterized by a high content of monosaccharides and disaccharides. In terms of diabetes prevention, evidence for sugar from solid sources on an association remains inconsistent (30). Hence, evidence for diabetes risk increasing properties of single food groups is limited to processed meat and potatoes. Still, the replicated positive association could possibly be caused in part by other contributing food groups, specifically the group “sugar.” It is also possible that food groups such as cakes and cookies, vegetable oils, and tea have rather diluted detrimental effects of more harmful food groups.

The replication of the “Simple Spain” score resulted in no significant association with diabetes risk. Exclusions of single countries in a sensitivity analysis resulted in no major changes of the pooled estimate. The pattern was characterized by potatoes, legumes, bread, red meat, processed meat, eggs, vegetable oils, wine, spirits, pasta and rice. For the food groups potatoes, red and processed meat (as discussed earlier), eggs (31), bread (mainly white bread), and pasta and rice (32, 33), recent studies underline greater risk. Thus, evidence for the majority of contributing food groups to this dietary pattern supports diabetes risk-increasing properties. Still, such positive association could be counterbalanced by moderate alcohol consumption, which is inversely associated with diabetes risk (34). However, the exclusion of wine from the “Simple Spain” score in a sensitivity analysis did not result in a stronger positive association.

The diabetes-association of the “Replicative France” score, characterized by contributions of nuts, “other fruits,” processed meat, fish, eggs, cake and cookies, coffee, and “other alcoholic beverages,” could not be replicated. For some components, evidence for an inverse association with diabetes risk is plausible, for example, nuts (34) and coffee (36), whereas processed meat and eggs were discussed before as diabetes risk factors. Evidence remains inconclusive in terms of intake of fish (37), whereas for “other fruits” evidence for no influence on diabetes risk is probable (38). “Other alcoholic beverages,” including fortified wine, could potentially explain the inverse association in France, as adjusting for this food group resulted in an attenuated association (data not shown). We observed a markedly higher intake of this food group in the female French study population in comparison to other countries. In a previous EPIC-InterAct publication, moderate consumption of fortified wine, compared to low intake, was associated with a 32% lower risk of diabetes in women (39). As the intake was markedly lower in other countries, exposure differences for this beverage might have been too low to affect diabetes risk.

The “Replicative Norfolk” score, identified as the only pattern to be significantly associated with an increased diabetes risk, showed rather limited comparability in pattern composition with other exploratory patterns previously described in European studies (8, 9). A comparison with patterns from the United States resulted in overlaps for potatoes (including French fries), processed meat, sugar, and cake and cookies, but tea and vegetable oils remain specific for the “Replicative Norfolk” score (40–42). Taking into account the aim of exploratory methods to deduce dietary patterns explaining maximum variance in food groups, but not necessarily being associated with the outcome of interest, limited comparability was expected. This was also confirmed in a recent systematic literature review, where, despite few exploratory dietary patterns with partially overlapping food groups, many population-specific dietary patterns remained incomparable (7). Still, among these, some were associated with diabetes risk as well (8, 43, 44).

The “Simple Spain” score, by contrast, overlapped in several food groups (refined grains, eggs, red meat, processed meat, and potatoes including French fries), identified in diabetes-associated patterns from single-country studies (7).

To discuss strengths and limitations of this analysis, the grouping of food items for PCA and the assumption of a set of foods to constitute all relevant food items remains difficult to justify (45). However, attempts were made in EPIC-InterAct to harmonize food groups and facilitate their use in dietary pattern analyses. Nevertheless, we could not consider different preparation techniques or the influence of meal timing. Hence, we still assume country-specificity in the resulting dietary patterns (11). Furthermore, the overall variance explained in the food groups ranged between 7.5% and 9.7%. Hence, these identified diabetes-associated dietary patterns did not reflect a high proportion of variance in the food groups. To minimize subjective decisions in PCA, we applied the eigenvalue ≥1 criterion which is straightforward. However, the second criterion of an identifiable turning point in the scree plots to separate principal components explaining relatively large proportions of variance in the food groups, may be ambiguous. Consequently, an additional criterion was consistently applied: the interpretability criterion, defined as cutoff for the number of food groups (≥3) and their factor loadings (≥0.4) (20). Pattern scores reflect relative, rather than absolute, quantitative differences in intake in comparison to a population average.
Although the initial derivation of patterns was based on the food intake distribution of the respective country, we used the full subcohort distribution for replication purposes. If intake distributions between country-specific subcohorts and the full subcohort differed greatly, effect estimates in the meta-analysis could have limited comparability to the country-specific estimates, therefore limiting the internal validity. Furthermore, the replication of only those dietary patterns being significantly associated with diabetes risk could be considered as a restrictive approach and it cannot be ruled out that other than the 3 identified country-specific patterns would show meaningful associations, if evaluated across all EPIC-InterAct countries. However, a lack of statistical significance for other patterns was mostly accompanied by weak associations, which supports our procedure. As strength of the replication, different cutoffs for factor loadings were systematically applied to investigate the capability of the simplified sum scores to reflect the original dietary pattern. The capability was quantified as Spearman correlation coefficient of \( r \geq 0.90 \), an arbitrary cutoff but ensuring sufficient reflection of the dietary pattern. As the Spanish dietary pattern was the only successfully simplified pattern, the replication of the other 2 diabetes-related patterns was done by applying the original pattern structure. However, this specific approach was feasible in the study setting of EPIC-InterAct with harmonized food groups. In other study settings, replication of original pattern structures may be strongly limited by lacking concordant food items and food groups. Although the use of simplified pattern scores appears to be more suitable in such cases, investigators should be aware that simplified sum scores may not sufficiently reflect the original pattern scores.

To conclude from this analysis, of the 18 dietary patterns derived by PCA in EPIC-InterAct, only 3 patterns were significantly associated with risk of type 2 diabetes in country-specific analyses. Of these, the UK-Norfolk pattern, characterized by high intakes of potatoes, processed meat, vegetable oils, sugar, cake and cookies, and tea, was largely consistently associated with increased diabetes risk across 8 countries. A Spanish pattern may also be related to increased risk, although associations were more heterogeneous across countries. Still, the UK-Norfolk and the Spanish dietary patterns show some similarity in food group composition (potatoes, processed meat), which are also characteristic for previously reported diabetes-related exploratory dietary patterns. Our study furthermore demonstrates that replication of associations of exploratory patterns with health outcomes is feasible and a necessary step to focus on generalizable associations from such analyses.

Acknowledgments

For information on how to submit an application for gaining access to EPIC data and/or biospecimens, please follow the instructions at [https://epic.iarc.fr/access/index.php](https://epic.iarc.fr/access/index.php). OM would like thank the participants of the Spanish EPIC cohort for their contribution to the study as well as to the team of trained nurses who participated in the recruitment. EPIC Ragusa acknowledges for their participation blood donors of AVIS RAGUSA (local blood donors association). We thank all EPIC participants and staff for their contribution to the study. We thank Nicola Kerrison (MRC Epidemiology Unit, Cambridge) for managing the data for the InterAct Project. The authors’ responsibilities were as follows—FJ: had access to all data for this study and takes responsibility for the manuscript contents; FJ: analyzed the data; MBS, and JK: contributed to the conception and design of the study, interpretation of the data, and drafted the manuscript; CL, SJS, NGF, NDK, and NJW: contributed to the conception and design of the study, interpretation of the data and critical revision of the manuscript; MJG, GF, HB, AT, KO, JRGQ, MJS, AB, NJW, TK, ER, DP, RT, CS, SP, AMWS, PN, OR, and PWF: contributed to the data collection, interpretation, and critical revision of the manuscript; CA, VC, SC-Y, CCD, CD, HF, TJK, KTK, CK, FRM, OM, MSS, RS, IS, and TYNT: contributed to the interpretation of the data and critical revision of the article for important intellectual content; and all authors: read and approved the final paper.
Exploratory dietary patterns and type 2 diabetes