Methodological approaches to compile and validate a food composition database for methyl-group carriers in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study

Heleen VAN PUYVELDE[a,b#], Vickà VERSELE[c#], Marlène DE BACKER[d], Corinne CASAGRANDE[e], Genevieve NICOLAS[e], Joanna L. CLASEN[f], Cristina JULIÁN[g], Guri SKEIE[h,i], Maria-Dolores CHIRLAQUE [j,k], Yahya MAHAMAT-SALEH[l,m], Pilar AMIANO ETXEZARRETA[n], Sara PAUWELS[o,p], Lode GODDERIS[o,q], Marc J. GUNTER[e], Koen VAN HERCK[a], Inge HUYBRECHTS[e]*, on behalf of the EPIC collaborators

Affiliations:

[#]Shared first authorship

[a]Department of Public Health and Primary Care, Faculty of Medicine and Health Sciences, Ghent University, C. Heymanslaan 10, K3, 9000 Ghent, Belgium

[b]Research Foundation - Flanders (FWO), Egmontstraat 5, 1000 Brussels, Belgium

[c]Department of Movement and Sport Sciences, Faculty of Physical Education and Physiotherapy, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

[d] Department of Nutrition and Dietetics, Faculty of Education, Health and Social Work, HoGent, Keramiekstraat 80, 9000 Gent, Belgium

[e]Nutrition and Metabolism Section (NME), International Agency for Research on Cancer, 150 cours Albert Thomas, 69372 Lyon CEDEX 08, France

[f] Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, St Mary's Hospital, Praed St, London W2 1NY, UK

[g]GENUD (Growth, Exercise, NUtrition and Development) research group, Department of Physiatry and Nursery, University of Zaragoza, Pedro Cerbuna, 12, 50009, Zaragoza, Spain

[h]Department of Community Medicine, UiT the Arctic University of Norway, Hansine Hansens veg 18, 9073 Tromsø, Norway

[i]Nutritional Epidemiology Group, School of Food Science and Nutrition, University of Leeds, Woodhouse Ln, Leeds LS2 9JT, UK

[j]Department of Epidemiology, Regional Health Council, IMIB-Arrixaca, Murcia University, Ronda de Levante, 11. E30008 Murcia, Spain

[k]CIBER in Epidemiology and Public Health (CIBERESP), Av. Monforte de Lemos, 3-5. Pabellón 11. Planta 0 28029 Madrid, Spain

[l]CESP, Fac. de médecine, Univ. Paris-Sud, Fac. de médecine - UVSQ, INSERM, Université Paris-Saclay, 114, rue Edouard-Vaillant, 94805 Villejuif, France

[m]Institut Gustave Roussy, 114, rue Edouard-Vaillant, 94 805 Villejuif, France

[n]Unidad Vigilancia Epidemiológica, Subdirección de Salud Pública de Gipuzkoa, Epic-Project San Sebastian, Public Health Division of Gipuzkoa, Nafarro hiribidea 4, 20013 San Sebastian, Spain

[o]Department of Public Health and Primary Care, KU Leuven, Kapucijnenvoer 35 blok D box 7001, 3000 Leuven, Belgium

[p]Unit Environmental Risk and Health, Flemish Institute of Technological Research (VITO), Vlasmeer 7, 2400 Mol, Belgium

[q]IDEWE (Externe dienst voor Preventie en Bescherming op het Werk), Interleuvenlaan 58, 3001 Heverlee

*Corresponding author: Inge Huybrechts, PhD. *Tel*: +33 (0) 4 72 73 81 48

Email addresses:
Heleen VAN PUYVELDE: heleen.vanpuyvelde@ugent.be
Vickà VERSELE: vicka.versele@vub.be
Marlène DE BACKER: debacker_marlene@hotmail.com

Corinne CASAGRANDE: casagrandec@iarc.fr

Genevieve NICOLAS: nicolasg@iarc.fr

Joanna L. CLASEN: j.clasen18@imperial.ac.uk

Cristina JULIÁN: cjulian@unizar.es

Maria-DOLORES CHIRLAQUE: mdolores.chirlaque@carm.es

Guri SKEIE: guri.skeie@uit.no

Yahya MAHAMAT-SALEH: Yahya.MAHAMAT-SALEH@gustaveroussy.fr

Pilar AMIANO ETXEZARRETA: epicss-san@euskadi.eus

Sara PAUWELS: sara.pauwels@kuleuven.be

Lode GODDERIS: lode.godderis@kuleuven.be

Koen VAN HERCK: koen.vanherck@ugent.be

Marc J. GUNTER: gunterm@iarc.fr

Inge HUYBRECHTS: huybrechtsi@iarc.fr

Abstract

A standardised methodology was used to compile and validate a methyl-group carrier database (MGDB) including folate, choline, betaine and methionine, for use in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. Compilation was performed by following structured guidelines to match the EPIC dietary intake data to food items from four food composition databases, according to their assigned priority of use. To assess relative validity, calculated dietary folate intakes were compared between the MGDB and the EPIC nutrient database (ENDB), used as the reference database. Folate intakes based on the MGDB and those generated using the ENDB showed good agreement (weighted κ = 0.63) and were strongly correlated (r = 0.81);

This MGDB can be used for investigating potential associations between methyl-group carrier intakes and risk or prognosis of cancer and other diseases in the EPIC study population.

Keywords: food composition database; methyl-group carriers; folate; choline; betaine; methionine; comparative study

Chemical compounds studied in this article

Folate (PubChem CID: 135398658); Choline (PubChem CID: 305); Betaine (PubChem CID: 247); Methionine (PubChem CID: 6137)

1. Introduction

Methyl-group carriers are nutrients such as folate, choline, betaine and methionine that carry a one-carbon (1C) unit which can be activated and transferred within a metabolic process, a mechanism known as 1C metabolism (Ducker & Rabinowitz, 2017). The methyl-group carriers enter 1C metabolism at different points, but all serve as precursors to S-adenosylmethionine (SAM) (Figure 1) (Anderson, Sant, & Dolinoy, 2012; Feil & Fraga, 2012). SAM, considered the universal methyl donor, supplies a 1C unit in methylation reactions, in.cluding DNA methylation (S. Friso, Udali, De Santis, & Choi, 2017).

Figure 1: Simplified illustration of one-carbon metabolism.

Dark blue: Methyl-group carriers; light blue: nutrients acting as coenzymes; white: intermediates within the 1C metabolism

Abbreviations: DHF dihydrofolate: ; THF: tetrahydrofolate; Vit B6: vitamin B6; Vit B2: vitamin B2; Vit B12: vitamin B12; DMG: dimethylglycine; SAM: S-adenosylmethionine; SAH: S-adenosylhomocysteine

DNA methylation has been suggested as an underlying molecular mechanism contributing to the effects of dietary factors on the development and progression of several diseases, including cancer (Jiménez-Chillarón et al., 2012). DNA methylation is a dynamic and potentially reversible process in which methyl-groups bind to the dinucleotides without changing the DNA sequence itself (Bird, 2002; Simonetta Friso & Choi, 2002). Modifications in DNA methylation patterns can affect gene expression or influence genome stability, leading to alterations in disease risk (Jiménez-Chillarón et al., 2012; Nazki, Sameer, & Ganaie, 2014).

Because of their presumed impact on DNA methylation through 1C metabolism, much attention has been given to methyl-group carriers in the diet. Deficient or excessive dietary intakes of methyl-group carriers might affect the availability of SAM and subsequently influence DNA methylation patterns and thus also cancer risk (McKay & Mathers, 2011). Research has begun to elucidate the effects of methyl-group carriers, folate and methionine in particular, on cancer risk; however, results are not robust. Adequate dietary intakes, before the appearance of preneoplastic tissue, potentially prevents tumour development (Chen, Li, Li, Li, Chu, & Wang, 2014; Wu, Cheng, & Lu, 2013), but overconsumption may contribute to the proliferation of already-initiated tumour cells (Cavuoto & Fenech, 2012; Cellarier et al., 2003; Ulrich, 2007).

Analyses in large-scale cohort studies investigating the role of dietary methyl-group carriers in 1C metabolism, DNA-methylation and associated disease outcomes are still scarce due to the lack of high-quality data on dietary methyl-group carriers. Detailed information on the chemical composition and nutrient yield of foods, based on chemical analysis can be found in food composition databases (FCDBs) (EuroFIR, 2020). In 1999, a study comparing nutrients in the FCDBs from nine European countries concluded that only France, The Netherlands and the United Kingdom (UK) provided FCDBs including comparable, methodologically correct folate values; the incomparable values resulted primarily from problems in the standard methods used and lack of clarity in the terminology and definitions (Deharveng, Charrondiere, Slimani, Southgate, & Riboli, 1999). In 2011, a critical evaluation of folate data in 15 European and three international FCDBs also stated a lack of comparability, mainly due to a lack of value documentation (e.g. method of measurement) and the use of generic terminologies (Bouckaert et al., 2011). Aside from folate, most of the European national FCDBs are lacking data on methyl-group carriers: none of them include choline or betaine, and only the German and Danish FCDBs contain methionine. Therefore, data from foreign FCDBs need to be used when assigning nutritional values of methyl-group carriers to dietary intake data. In order to evaluate methyl-group carrier intakes and their associations with adverse health outcomes such as cancer, a standardised FCDB for folate, choline, betaine, and methionine is needed for use in the European Prospective Investigation into Cancer and Nutrition (EPIC) study.

This paper aims to describe the methodology used to compile a methyl-group carrier database (MGDB) for epidemiological research, using four foreign FCDBs and dietary assessment data from the EPIC study. In addition, this project allows for the assessment of the overall quality of the applied methodology by examining the comparability of the dietary folate intakes determined by two different approaches: a) this more pragmatic approach to compile a MGDB using four available FCDBs and b) a similarly standardised approach preferentially using national FCDBs (Nicolas et al., 2016).

2. Materials and methods

2.1. EPIC Study design

Briefly, the EPIC study is an ongoing prospective cohort study aiming to investigate the role of dietary habits and nutritional status, as well as a wide range of environmental and lifestyle factors in relation to cancer and disease morbidity (Riboli et al., 2002; Riboli & Kaaks, 1997). Between 1992 and 2000, this project enrolled 521,324 apparently healthy men and women (age 20 - 84 years) from 23 recruitment centres across ten European countries (Denmark, France, Germany, Greece, Italy, Norway, Spain, Sweden, the Netherlands, and the UK) (Riboli et al., 2002). The rationale, design and methods of the EPIC study have been described elsewhere (Riboli et al., 2002). The ethical review boards of the International Agency for Research on Cancer (IARC – Lyon, France) and those of all participating recruitment centres approved the EPIC study. Written informed consent was provided by all EPIC participants in order to process their data.

2.2. Dietary assessment within EPIC

Within the EPIC study, the prospective cohort approach included the collection of information at baseline through country-specific, validated dietary questionnaires (DQ), designed to capture individual long-term usual dietary intake and geographical specificity of the diet (Riboli et al., 2002). To calibrate dietary intake measurements obtained through these different DQ, a computer-assisted, single 24-hour dietary recall (24-HDR) interview program (EPIC-soft) was used by trained interviewers (Slimani, Ferrari, Ocke, & Welch, 2000). The program was designed to conduct interactive, by telephone (Norway) or face-to-face dietary interviews according to a procedure that was standardised within and between EPIC centres (Slimani et al., 2000). The 24-HDR was collected in a representative sample (N =36,994) of the entire EPIC cohort (Slimani et al., 2002).

2.3. Initial compilation of a harmonised nutrient database for the EPIC project

The EPIC Nutrient Database (ENDB), which originally focused on 26 priority components, was compiled at the end of the nineties to harmonise the nutrient values of national FCDBs across the 10 participating EPIC countries (Slimani et al., 2007). Methyl-group carriers were not included during the ENDB-project due to the absence of (comparable) food composition data on methyl-group carriers across FCDBs in the different EPIC countries (Deharveng et al., 1999). Since 2010, a folate database has been compiled as an extension of the ENDB, based on a new inventory focused on folates (Bouckaert et al., 2011). Nutrient values, preferentially obtained from the national FCDBs of the respective EPIC countries were adopted, using standardised procedures. The in-depth process for compiling this EPIC folate database was described elsewhere (Nicolas et al., 2016).

2.4. Selecting food composition data sources for methyl-group carriers

To date, none of the national FCDBs of the ten EPIC countries contain methodologically reliable nutritional values for all four methyl-group carriers: folate, choline, betaine, and methionine. Standard reference analytical methods are microbiological assay (MA) for folate (Greenfield & Southgate, 2003), liquid chromatography-electrospray ionization-isotope dilution mass spectrometry for choline and betaine (Koc, Mar, Ranasinghe, Swenberg, & Zeisel, 2002), and performic oxidation/ high performance liquid chromatography (HPLC) for methionine (Greenfield et al., 2003). In the past few years, these methyl-group carriers have been incorporated into a few

FCDBs including the U.S. FCDB (National Nutrient Database for Standard Reference of the U.S. Department of Agriculture - USDA) and the Canadian FCDB (Canadian Nutrient File). Both of these FCDBs include all four nutrients of interest, a large number of food items and made use of the standard reference analytical methods. Betaine and choline were only included in the U.S. FCDB since 2008. Two European databases include nutritional data concerning methyl-group carriers other than folate, obtained by the reference analytical methods: the Danish FCDB (Danish Food Composition Databank) and the German FCDB (Bundeslebensmittelschlüssel), which include methionine as well as folate.

In order of priority, the U.S. FCDB, Canadian FCDB, German FCDB, and Danish FCDB were used to compile the MGDB for EPIC. Priority was determined based on the quality of the analytical methods used, the availability of the maximum number of methyl-group carriers and the total number of food items comprising nutritional values of the respective methyl-group carriers. Compilation of this MGDB took place between 2014 and 2017. Further details on the four FCDBs used for this compilation are listed in Appendix 1.

2.5. Food composition database compilation

The compilation of the MGDB builds on the procedure of the aforementioned folate database of the ENDB (Nicolas et al., 2016), which is based on the general concepts of the original ENDB project (Slimani et al., 2007). The matching was first performed for the food items derived from the 24-HDR data (Figure 2). Subsequently, links between food items reported in the 24-HDR and DQ, set during the ENDB project, were used to assign nutrient values to DQ food items. DQ items with no link with 24-HDR items were matched using the U.S. FCDB exclusively, following the same procedure as described in Figure 2.

Figure 2: The compilation process of the methyl-group carrier database (MGDB)

Consumed foods reported in the EPIC 24-HDR were described in a detailed and systematic way. Therefore, the food list from the EPIC 24-HDR, rather than DQ data, was used as the starting point for the compilation of the nutrient database (Slimani et al., 2007). This resulted in a high number of different food items for each country and was reduced to bring it to the same level of detail as the foods provided in the FCDB, as described in detail by Slimani et al (2007) and Nicolas et al (2016). Briefly, food items were aggregated using common rules across countries and with respect to their relevance to cancer research. A total number of 547-1,537 food items per country were included in the final food list to compile the MGDB (Nicolas et al., 2016).

2.5.1. General guidelines for matching food items

The EPIC food items were linked to one of the food items available in the four FCDBs, taking into account their priority. If an exact match could be found, nutritional values for the respective methyl-group carriers were assigned directly. However, some specific food items (e.g., different types of cheese) could not be found in any of the four FCDBs used. In that case, the matching process included an equivalency check between the reported food items and similar food items available in the used FCDBs on the basis of their definition, description and nutritional composition as described in the ENDB (e.g. red Leicester cheese was linked to cheddar cheese).

Although the EPIC-Soft 24-HDR interview programme allowed for the collection of detailed and standardised data, some reported foods lacked sufficiently detailed descriptions or specifications to allow an exact or equivalent match. These food items were coded as ‘not specified’ (n.s.) and a weighted average based on the frequencies of consumption of equivalent reported foods was assigned (e.g., vegetable oil n.s.: weighted average of all vegetable oils including olive oil, rapeseed oil, corn oil, etc.). These food items were named ‘generic items’.

Nutritional values for multi-ingredient foods (composite foods in particular, e.g. béarnaise sauce, mango chutney or fruit scones) which were not available in any of the FCDBs were obtained by recipe calculations, considering the use of retention factors (corrects for changes in the nutrient composition of food by thermal processing) at the ingredient level and yield factor (corrects for weight changes due to food preparation methods) at the recipe level, if relevant. The existing country-specific recipe files of the ENDB project, provided by the EPIC partners, were used as recipe sources. If no suitable recipe was found, a new recipe was created by breaking down the composite foods into their single, least modified ingredients. The single ingredients were treated as separate food items to match with the FCDBs, and were consequently subject to recipe calculations.

In case no exact or equivalent match could be found for a single food item or ingredient, nutritional values for methyl-group carriers were obtained by applying different available algorithms, yield factors and retention factors, depending on the nature of the food item. This included calculation methods to adjust for raw-to-cooked water losses/gains and mineral and vitamin losses of the FCDB item. These approaches were mainly applicable for single food items (e.g. fat-reduced cheese), or single foods cooked using cooking methods not available in the four selected FCDBs. Food items subject to these algorithms were called ‘one-ingredient recipes’.

2.5.2. Guidelines for matching food items: special cases

To properly match foods with different cooking methods to food items in the four FCDBs, the same rules for food linkage as used in the ENDB project were applied (Slimani et al., 2007). Foods cooked without fat (e.g. boiled or steamed) were preferably matched to an exact or similar cooked food item in the FCDBs. In case an exact or similar match was not possible, the food item was treated as a one-ingredient recipe by matching the cooked food item to its raw variant and applying the calculation methods described in paragraph 2.5.1. On the other hand, foods cooked with fat were systematically treated as two separate food items: the raw food and its specified fat. Both food items had to be linked to the FCDBs and subsequently adjusted for cooking using the algorithms, yield and retention factors.

Likewise, canned food items were preferentially linked to an identical drained canned item. However, a canned item was considered similar to a boiled/steamed item when no exact match could be found in the FCDBs. Frozen items were linked to raw items if no frozen item was available. Priority was given to the least modified food item if no information on the state of processing was specified (e.g., “cooked without salt” was chosen over “cooked with salt”, and “vegetables with skin” were prioritised). No fortified food items were included in the MGDB, unless the food item was described as enriched with folate.

2.5.3. Additional efforts to complete the database

To limit missing values, logical zero values for methionine were assigned to all foods containing no protein. For betaine and choline, logical zero values were assigned to products such as water and artificial sweeteners. Thereafter, all remaining missing values were replaced by zeros to allow the calculation of methyl-group carrier intakes for all subjects in further analyses.

Two quality controls were performed to guarantee the accuracy of the food matching and avoid errors. First, blinded re-matching of a random sample of food items was performed independently by two researchers. Second, the fully completed files were checked twice: once by an accredited nutritionist and once by an expert of the ENDB project.

Although country-specific folate values had already been included in the ENDB, alternative values were derived using the four selected FCDBs. This created the opportunity to carry out comparative analyses between our approach and the folate ENDB approach in which all EPIC countries used preferably local FCDBs, completed with other FCDBs such as the U.S. FCDB when local data were missing (Nicolas et al., 2016).

2.6. Statistical Analyses

Reported food intakes from participants of the 24-HDR and the DQ were analysed in this study. To reduce the impact of outliers, participants at the lowest and highest 1% of the distribution of the ratio of reported total energy intake to energy requirement were excluded from the analyses for the DQ data. No exclusions were carried out regarding the data of the 24-HDR because of its detailed and standardised nature and built-in quality controls.

Descriptive analyses were carried out to report missing values for folate, choline, betaine and methionine (before replacement by logical zeros). To evaluate the relative validity of the newly compiled MGDB, dietary folate intakes calculated by the MGDB were compared to dietary folate intakes calculated by the ENDB, used as the reference database in this study. Therefore, absolute and relative differences in dietary folate intakes were examined. Relative measurements are of great importance because accurate ranking and categorising of individuals according to their dietary intakes is the main requirement for further epidemiological analyses.

To report on absolute differences in dietary folate intakes obtained by the ENDB and the MGDB, mean differences were calculated using the method proposed by Giavarina (2015), and paired samples t-tests were carried out, both globally and stratified by the ten EPIC countries involved. Relative differences in dietary folate intake between the ENDB and the MGDB were examined using Pearson correlations, Bland-Altman plots and weighted kappas. Pearson correlation coefficients were calculated to assess the associations between dietary folate intakes estimated using the ENDB and the MGDB. To further investigate the agreement between these methods, a Bland-Altman test was used (Bland & Altman, 1986), presented as mean difference percentage plots and the corresponding limits of agreement within which an estimated 95% of the differences in dietary folate intake fall (Giavarina, 2015). For the Bland-Altman plots, differences in folate intakes between the databases (displayed on the y-axis) were expressed as percentages as there is an increase in variability of the differences with increasing magnitude of the mean folate intakes (Giavarina, 2015). The agreement of the classification of individual folate intakes into quintiles was calculated and tested by weighted kappa coefficients. Cut-offs for quintiles were assigned separately for the two databases.

Non-parametric tests (Spearman correlations and Wilcoxon signed-rank tests) were performed as a sensitivity analysis. As results were very similar, only results of the parametric tests were reported.

All statistical tests were carried out for the 24-HDR data and DQ data as two-sided tests and with a statistical significance level of $\alpha = 0.05$. Statistical analyses were carried out with the Statistical Package for the Social Sciences (SPSS Inc., Chicago, IL, USA) version 20.0.

3. Results

A description of the matched food items is shown in Table 1, for both the 24-HDR and the DQ food data. Regarding the 24-HDR data, a total of 10,173 food items were included for matching, of which 5,069 (49.8%) were categorised as an exact or equivalent match. For 4,926 food items (48.4%), recipes were applied to compute the nutritional values - including ‘one-ingredient recipes’. The remaining food items (N =178; 1.7%) were generic items. Concerning the DQ data, 13,951 food items had to be matched, of which 9,692 (69.5%) were an exact or equivalent match, 1,796 (12.9%) food items were treated as a ‘recipe’ or ‘one-ingredient recipe’ and 2,463 (17.6%) food items were deemed generic items.

For the 24-HDR data, the U.S. FCDB was responsible for 87.1% of all exact or equivalent matches made, followed by the Danish FCDB (5.2%), the Canadian FCDB (4.3%) and the German FCDB (3.3%). For the DQ data, the U.S. FCDB had a much larger share (97.4%), followed by the Danish

FCDB (1.2%), German FCDB (0.8%) and Canadian FCDB (0.6%) to obtain the exact or equivalent matches.

The distribution of missing values for folate, choline, betaine and methionine for the exact matches in the MGDB is shown in Table 1. In both the 24-HDR and DQ data, the lowest number of missing values was found for folate (1.8% and 1.9% respectively) and the highest number was found for betaine (48.8% and 46.3% respectively).

Table 1: Description of the matched food items and the number of missing values for methyl-group carriers in the MGDB

	24-HDR N (%)	**DQ** N (%)
Food items (total)	10,173	13,951
Food items treated as:		
Generic items	178 (1.7%)	2,463 (17.6 %)
(One-ingredient) recipes	4,926 (48.4 %)	1,796 (12.9 %)
Exact match	5,069 (49.8%)	9,692 (69.5 %)
Food items matched to (exact matches only):		
U.S. FCDB	4,417 (87.1%)	9,437 (97.4 %)
Canadian FCDB	168 (4.3 %)	63 (0.6 %)
Danish FCDB	265 (5.2 %)	114 (1.2 %)
German FCDB	219 (3.3 %)	78 (0.8 %)
Missing values (exact matches only):		
Folate - ENDB	0 (0.0%)	54 (0.4%)
Folate - MGDB	178 (1.8%)	259 (1.9%)
Choline - MGDB	1,790 (17.6%)	1,951 (14.0%)
Betaine - MGDB	4,969 (48.8%)	6,458 (46.3%)
Methionine - MGDB	1,292 (12.7%)	1,646 (11.8%)

Abbreviations: MGDB: methyl-group carrier database; ENDB: EPIC nutrient database; 24-HDR: 24-hour dietary recall; DQ: dietary questionnaire; N: number

Reported food intakes of 36,994 participants for the 24-HDR data and 504,245 participants for the DQ data were analysed in this study. Table 2 shows the differences in mean dietary folate intakes between the ENDB and the MGDB. Results by country can be found in Appendix 2. For both the 24-HDR and DQ data, estimated dietary folate intakes were higher when calculated by the new MGDB procedures (24-HDR: 325.91 µg/day, SD =159.30; DQ: 354.56 µg/day, SD =127.84) compared to the ENDB (24-HDR: 265.25 µg/day, SD =137.83; DQ: 308.55 µg/day, SD =120.14). All stratified analyses showed this trend except for the DQ data in the UK, which had slightly, but still significantly, lower folate intake reported for the MGDB (396.17 µg/day; SD =129.26) compared to the reference ENDB (408.76 µg/day; SD =157.68). Italy, Spain and Germany showed the highest numbers of significant differences of the mean folate intakes between the approaches.

Table 2: Paired sample t-tests and mean differences for individual dietary folate intake between the compiled MGDB and the ENDB

	Folate (µg)	N	Mean (µg/day)	SD	Mean ∆ (µg/day)*
24-HDR data	MGDB	36,994	325.91	159.30	-60.66$^{\#}$ (-20%)
	ENDB	36,994	265.25	137.83	
DQ data	MGDB	504,247	354.56	127.84	-46.01$^{\#}$ (-14%)
	ENDB	504,247	308.55	120.14	

* Mean difference (%):The MGDB mean minus the ENDB mean (divided by their arithmetic mean [*100%])

$^{\#}$ Statistical difference p < 0.001 for the paired sample t-test

Abbreviations: N: number; SD: standard difference; Mean ∆: mean difference; 24-HDR: 24-hour dietary recall; DQ: dietary questionnaire; MGDB: methyl-group carrier database; ENDB: EPIC nutrient database

Strong correlations for dietary folate intakes were shown between the ENDB and the MGDB for both the 24-HDR data (r =0.73; p <0.001) and the DQ data (r =0.81; p <0.001). Results per country can be found in Appendix 3. Bland-Altman plots for the 24-HDR data and DQ data are presented in Figure 3. The mean difference, or bias, for the 24-HDR was -20.26% (SD = 29.80%) and the limits of agreement ranged from -78.66% to 38.14% (mean difference ± 1.96*SD). Concerning the DQ-data, a bias of -14.31% (SD = 19.48%) was found with limits of agreement ranging from -52.48% to 23.87%.

Figure 3: Bland-Altman plots for a) 24-HDR data and b) DQ data representing the mean differences of folate intake (in percentages) between the reference ENDB and the MGDB and their limits of agreement.

Legend: full line: mean difference in folate intake (%) calculated as the ENDB mean minus MGDB mean divided by their arithmetic mean (*100%); dotted line: limits of agreement (%) calculated as the mean difference in folate ± 1.96*SD (*100%);

The proportion of the participants classified into the same quintile for folate intake according to the reference ENDB and the newly created MGDB is 46% and 50% for the 24-HDR data and DQ data respectively (Table 3). If adjacent quintiles are also included, this increases to 86% (24-HDR data) and 91% (DQ data). Of all participants, 0.28% and 0.04% for respectively the 24-HDR data and DQ data were misclassified into the extreme opposite quintile. Results of the weighted kappa analysis indicated moderate agreement (weighted κ =0.56) in case of the 24-HDR data and good agreement (weighted κ =0.63) according to the DQ data, for folate intakes. Results per country can be found in Appendix 4.

Table 3: Weighted Kappas for individual dietary folate intake between the compiled MGDB and the ENDB

	Classified into the same Q (%)	**Classified into the adjacent Q (%)**	**Weighted κ**	**SE**	**CI lower**	**CI upper**
24-HDR data	46.17	39.83	0.56	0.003	0.55	0.56

DQ data	50.59	40.53	0.63	0.001	0.63	0.63

Abbreviations: MGDB : methyl-group carrier database ; ENDB : EPIC nutrient database ; Q: quintile; κ: kappa; SE: standard error; CI: 95% confidence interval

4. Discussion

The aim of this project was to generate a MGDB for use in the EPIC study in order to further investigate the relationship between dietary intakes of methyl-group carriers and health and disease outcomes. Therefore, dietary data from the ten European countries participating in the EPIC study were matched with food items from four selected FCDBs (in order of priority: U.S. FCDB, Canadian FCDB, Danish FCDB and German FCDB), using standardised procedures based upon those developed in the ENDB project.

The majority of nutritional values for the methyl-group carriers were derived from the U.S. FCDB, completed with information from the three other databases. The larger share from the U.S. FCDB can be attributed to the order of priority that was defined among the selected FCDBs, based on the quality of the analytical methods used, the availability of all methyl-group carriers of interest, and the exhaustiveness of the food list. The U.S. FCDB and Canadian FCDB provided values for all four methyl-group carriers, while the German FCDB and Danish FCDB only provided values for folate and methionine. Additionally, all FCDBs except for the German FCDB contained missing nutritional values for certain food items which led to numerous missing values in the MGDB, particularly for betaine. Food compilers prioritize their laboratory analysis for most frequently consumed foods or for certain nutrients by giving priority to the foods that most likely contain the nutrient to be analysed (Haytowitz et al., 1996). Therefore, missing values appear more often for foods that only contain traces or none of the nutrients under study. As such, many of the missing values in FCDBs can be considered as logical zeros, meaning that the component is not expected in that particular food item. The ENDB showed no missing values for folate because any available folate data for a food item or from a similar food was accepted from neighbouring countries or from the U.S. FCDB when no values analysed by MA could be found (Nicolas et al., 2016).

Even though folate had already been included in the ENDB, a second linking of the food items was carried out using the four selected FCDBs. This created the opportunity to assess the relative validity of the food matching performed in this study, while using the ENDB folate values as a reference (Nicolas et al., 2016).

Comparative analyses showed differences between dietary folate intakes estimated by the ENDB and the MGDB for participants in the EPIC study. For the dietary assessment data derived from the 24-HDR and the DQ, calculated mean dietary folate intakes were higher using the new MGDB compared to the ENDB, except for the UK DQ data. A plausible explanation for this difference is the use of more recently updated FCDBs to compile the MGDB compared to the ENDB, meaning that nutritional values for methyl-group carriers measured by MA have been recently assigned to a larger amount of the FCDB's food items. MA may provide higher folate values than other analytical methods. Additionally, product reformulation should be taken into account when using more recently updated FCDBs, which is important because the food industry has a high turnover of products. Therefore, it would be preferable to match nutritional data from the same time period as the baseline dietary assessment, especially for processed foods and composite foods. However, as previously highlighted, methodologically correct folate data were too scarce at that time. Another possible explanation for the differences between dietary folate intakes is the use of preferably country-specific FCDBs to compile the ENDB compared to the use of mainly the U.S. FCDB for compilation of the MGDB. There is likely a variation, especially in the content of vitamins and minerals, between different samples of the same food used in the different FCDBs. These differences in food composition can be found between regions (e.g. European carrots versus American carrots), but differences are also likely to be found between foods originating from the same geographic region or even from the same grower or manufacturer (e.g. one carrot can be more exposed to sunlight or pesticides then another carrot growing on the same field). Taking also into account import and export of foods between regions, it is hard to conclude on real regional variation in food composition. This concern supports the selection of one or few high-quality FCDBs that meet our selection criteria, above the constrained use of merely country-specific FCDBs.

Regarding the comparability of the two databases, it should be noted that national FCDBs sometimes use foreign FCDBs as a source of folate values analysed by MA. In the ENDB, the number of reported folate values analysed by MA ranged from 43% - 70%. Within this subset, between 14% (UK, France) and 27% (Italy) of folate values were borrowed from the U.S. FCDB release 21 (Nicolas et al., 2016).

Because of a lack of national nutritional values for methyl-group carriers, the U.S., Canadian, German and Danish FCDBs were used to compile the MGDB. This created difficulties for finding an appropriate match for each food item. International comparisons are more complex since each country has unique typical and local foods and meals. Identification of these kinds of foods and meals might be difficult, and assigned values taken from similar foods may be unreliable. Another possible explanation for the difference in intakes could be fortification, whether or not done nationally, which can result in different folate content of the same food items in the two databases. However, no fortified food items were included in the MGDB, unless it was described as enriched with folate.

Although significant differences in mean values were reported, strong correlations were found between folate intakes, demonstrating a good ranking of the subjects according to their folate intake. Also, results of the weighted kappa analysis indicated moderate agreement for the 24-HDR (weighted κ =0.56) and good agreement for the DQ folate intakes (weighted κ =0.63). The agreement between folate intakes is at least satisfactory, as 86% (24-HDR) and 91% (DQ) of the participants are classified into the same or adjacent quintile. Furthermore, Bland-Altman plots indicated good agreement between dietary folate intakes. The average discrepancy between methods, or bias, was acceptable (-20.26% for 24-HDR data; -14.31% for DQ data). This small bias goes with rather narrow limits of agreement, within which an estimated 95% of the differences in dietary folate intake fall, indicating that the two methods are sufficiently similar. Results of the DQ show consistently higher agreement compared to results of the 24-HDR. This is most likely due to the fact that food items in the 24-HDR were described in more detail compared to the DQ. Therefore, the matching procedure was more complex for the 24-HDR which could lead to extra bias. These comparative folate analyses demonstrate good relative validity of the new MGDB for ranking and categorising individuals according to their folate intakes; the main requirement in epidemiological cohort studies.

Previous studies have compared nutrient intake data calculated via different procedures and by different FCDBs. One such study examined the level of agreement between macro- and micronutrients of the U.S. FCDB (modified by Chilean food items) and the British FCDB. High to excellent agreement was found for all macronutrients (intra-class correlation coefficient (ICC) ranged from 0.96 (95% CI: 0.95–0.98) for proteins to 0.98 (95% CI: 0.98–0.99) for total fat) and for vitamin A (ICC: 0.998, 95% CI: 0.995–1.00) and vitamin C (ICC 0.995, 95% CI: 0.992–0.998), respectively). However, the interpretation for other vitamins and especially minerals was more uncertain (Garcia, Rona, & Chinn, 2004). In most of the studies, comparisons were made between European FCDBs (Deharveng et al., 1999; Hakala, Knuts, Vuorinen, Hammar, & Becker, 2003; Julian-Almarcegui et al., 2016; Slimani et al., 2007; Vaask et al., 2004). The use of non-national FCDBs in these studies could be partially justified since strong correlations ($r > 0.70$) have been found between the different European FCDBs, but these correlations apply mostly for macronutrients (Deharveng et al., 1999; Hakala et al., 2003; Julian-Almarcegui et al., 2016). However, some comparative studies suggest a discrepancy between FCDBs (Vaask et al., 2004). Research has shown that some nutrients, mostly micronutrients, are not analysed and expressed in a compatible way between nutrient tables, resulting in values that are not always comparable (Deharveng et al., 1999; Hakala et al., 2003; Vaask et al., 2004). This issue favours the use of one or few high quality FCDBs above the use of very different and lower quality regional FCDBs for multi-centre cohorts that include countries with very different levels of food composition data availability and quality. Indeed, differences between FCDBs are often more due to differences in laboratory methods used rather than true differences in food composition between regions (Deharveng et al., 1999; Nicolas et al., 2016).

It has long been recognised that folate values are difficult to harmonise when comparing national FCDBs (Bouckaert et al., 2011; Deharveng et al., 1999). Concerning their comparability, extra attention should be given to the source of nutritional values (i.e. analytical methods used to measure the nutrient content of foods, calculations or published literature by the food industry), accuracy in the definitions of nutrients and unit of measurement (Leclercq, Valsta, & Turrini, 2001). Furthermore, folate is an unstable component as it is labile to temperature, pH and oxidation, leading to potential problems in the measurement of this nutrient (Deharveng et al., 1999).

Given the various arguments that can explain differences between FCDBs, it is reassuring that in this project a satisfactory level of agreement for folate intake between the ENDB and the MGDB was shown. However, the results of the relative validation study for folate might not be generalisable to the other methyl-group carriers, especially betaine, which showed considerably more missing values compared to choline or methionine. Frequent missing values may lead to underestimation of the true betaine intakes. Comparison with nutritional biomarkers could potentially further assess the validity of these methyl-group carrier estimates in the EPIC study; although endogenous mechanisms may mask expected correlations between intakes and blood levels. The lack of food composition data for several food items for betaine, and to a lesser extent also for choline and methionine, is a limitation of this study. It may affect exposure estimations (underestimation of true intakes) and lead to the attenuation of associations found between methyl-group carrier intakes and health outcomes. However, most missing values concern food items that are not frequently consumed or that contain only traces or none of the methyl-group carriers (Haytowitz et al., 1996). Therefore, the impact of missing values is likely to be minimal. Yet, this emphasizes the need for valid food composition data on the methyl-group carriers to estimate individual nutrient intakes in order to provide better epidemiological evidence on their associations with disease risk.

To the best of our knowledge, this study is the first to compile a database of methyl-group carriers other than folate for international use. Two major strengths should be highlighted. First, in order to optimise accuracy and continuity, a standard procedure was maintained, building on the previous experiences of the ENDB project. For example, calculation principles (e.g. algorithms and retention factors) between databases were standardised, and country-specific recipes and generic food weightings were used because there are differences in recipes and food preparation methods between countries. Secondly, two complimentary, comprehensive quality controls were performed during the matching procedure to assure a systematic and standardised linking. Furthermore, the compilation of a MGDB is a valuable addition to the EPIC study. The establishment of the estimated dietary methyl-group carrier intakes, as new variables to explore in the EPIC cohort, will provide researchers with the opportunity to investigate additional risk factors for specific cancers and other chronic diseases. This is in alignment with the increasing amount of existing evidence indicating the importance of the methyl-group carrier nutrients (Obeid, 2013; Wallace et al., 2018).

5. Conclusion

This project demonstrates the complexity of matching food consumption data from an international cohort with FCDBs from other regions. However, this pragmatic approach for matching dietary assessment data to foreign FCDBs compares adequately to the ENDB approach adopting nutrient values from national FCDBs of the EPIC countries. Therefore, this methodology for matching food items from multi-centre cohorts to one or a few high-quality FCDBs, has the potential to be a framework to build off for other similar projects. Strong correlations and moderate to good levels of agreements were shown for folate intakes. However, to date there are no resources available to examine to what extent this can be generalised to the other three methyl-group carriers, in particular for betaine. As there were many missing values for betaine, more efforts are needed to include comparable values across national FCDBs, using reference analytical methods for assessing the nutrient contents of the foods.

This methyl-group carrier intake data in EPIC will assist in disentangling the role of dietary methyl-group carriers in 1C metabolism, DNA methylation and disease risk.

Acknowledgments

The authors would like to thank the EPIC study participants and staff for their valuable contribution to this research. The authors would also like to thank Mr. Bertrand Hemon and Ms. Carine Biessy for their support in preparing the databases. H.V.P. was supported by a PhD fellowship of the Research Foundation Flanders (FWO, 189019N). The coordination of EPIC is financially supported by the European Commission (DG-SANCO); and the International Agency for Research on Cancer. The national cohorts are supported by Danish Cancer Society (Denmark); Ligue Contre le Cancer; Institut Gustave Roussy; Mutuelle Générale de l'Education Nationale; and Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), and Federal Ministry of Education and Research (BMBF) (Germany); Hellenic Health Foundation; Stavros Niarchos Foundation; and the Hellenic Ministry of Health and Social Solidarity (Greece); Italian Association for Research on Cancer (AIRC); National Research Council; and Associazione Iblea per la Ricerca Epidemiologica (AIRE-ONLUS) Ragusa, Associazione Volontari Italiani Sangu (AVIS) Ragusa, Sicilian Government (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS); Netherlands Cancer Registry (NKR); LK Research Funds; Dutch Prevention Funds; Dutch ZON (Zorg Onderzoek Nederland); World Cancer Research Fund (WCRF); and Statistics Netherlands (the Netherlands); and Nordic Center of Excellence Programme on Food, Nutrition and Health (Norway); Health Research Fund (FIS); Regional Governments of Andalucía, Asturias, Basque Country, Murcia (No. 6236) and Navarra; and the Centro de Investigación Biomédica en Red en Epidemiología y Salud Pública and Instituto de Salud Carlos II (ISCIII RETIC) (RD06/0020) (Spain); Swedish Cancer Society; Swedish Scientific Council; and Regional Government of Skåne and Västerbotten (Sweden); Cancer Research UK; Medical Research Council; Stroke Association; British Heart Foundation; Department of Health; Food Standards Agency; and the Wellcome Trust (UK). Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

IARC disclaimer: Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

Conflict of interest

The authors declare no conflict of interests in relation to the work described.

Figure 1: Simplified illustration of one-carbon metabolism.

Dark blue: Methyl-group carriers; light blue: nutrients acting as coenzymes; white: intermediates within the 1C metabolism

Abbreviations: DHF dihydrofolate: ; THF: tetrahydrofolate; Vit B6: vitamin B6; Vit B2: vitamin B2; Vit B12: vitamin B12; DMG: dimethylglycine; SAM: S-adenosylmethionine; SAH: S-adenosylhomocysteine

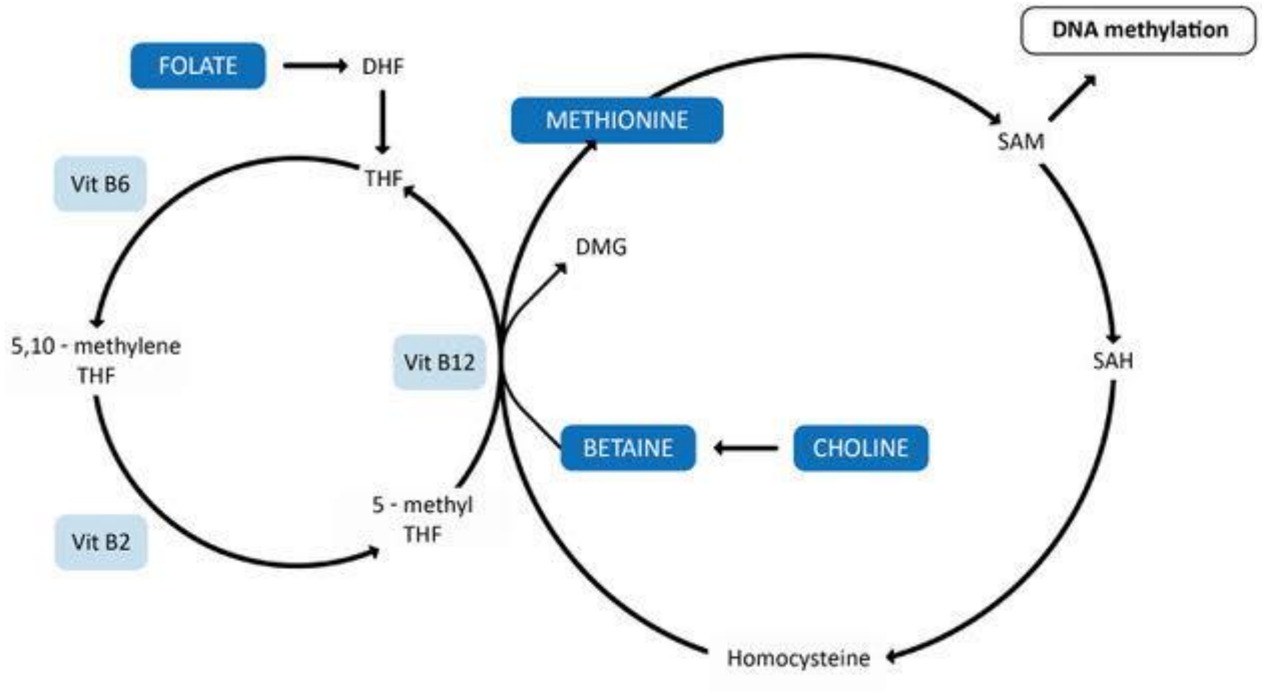


Figure 2: The compilation process of the methyl-group carrier database (MGDB)

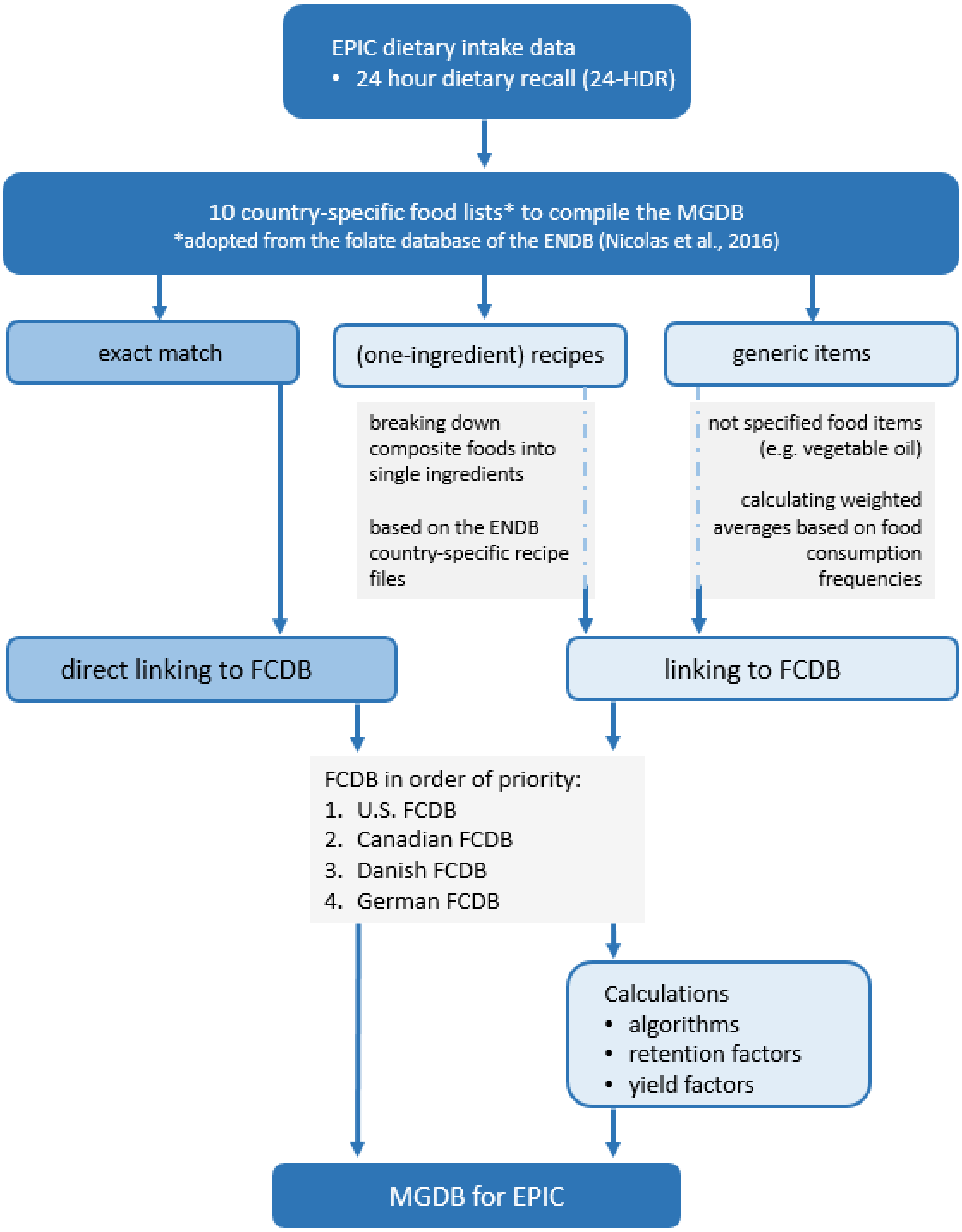
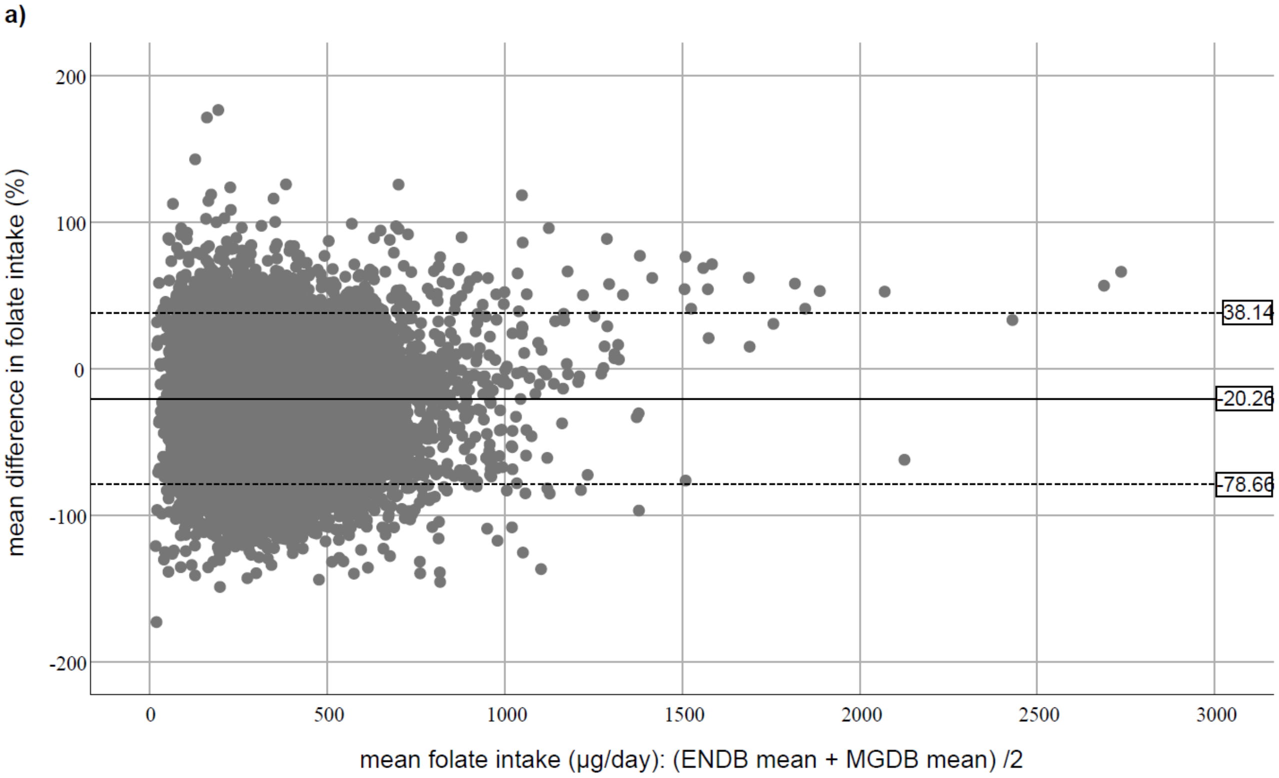
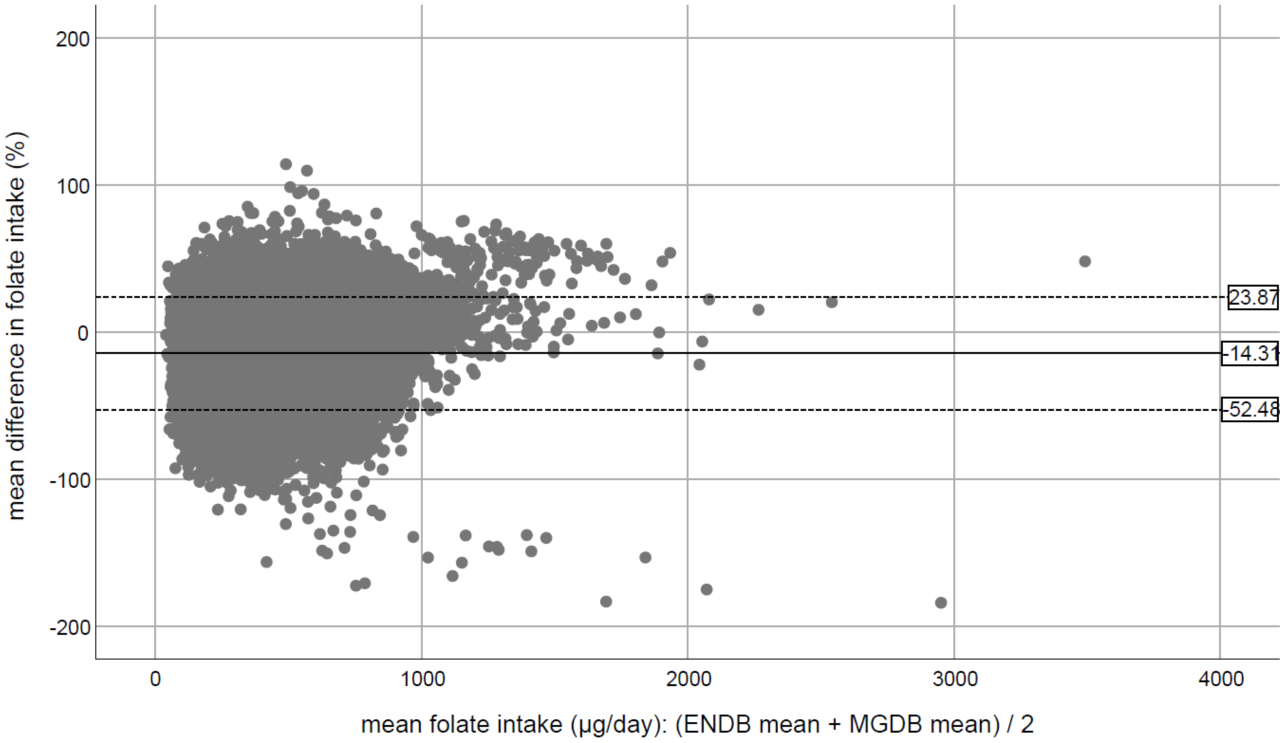


Figure 3: Bland-Altman plots for a) 24-HDR data and b) DQ data representing the mean differences of folate intake (in percentages) between the reference ENDB and the MGDB and their limits of agreement.

Legend: full line: mean difference in folate intake (%) calculated as the ENDB mean minus MGDB mean divided by their arithmetic mean (*100%); dotted line: limits of agreement (%) calculated as the mean difference in folate ± 1.96*SD (*100%);



b)


200
100
0
-100
-200
mean difference in folate intake (%)
23.87
-14.31
-52.48
0
1000
2000
3000
4000
mean folate intake (µg/day): (ENDB mean + MGDB mean) / 2

References

Anderson, O. S., Sant, K. E., & Dolinoy, D. C. (2012). Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation. *The Journal of nutritional biochemistry, 23*(8), 853-859. https://doi.org/10.1016/j.jnutbio.2012.03.003.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes & development, 16*(1), 6-21. https://doi.org/10.1101/gad.947102

Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet, 327*(8476), 307-310.

Bouckaert, K. P., Slimani, N., Nicolas, G., Vignat, J., Wright, A. J., Roe, M., . . . Finglas, P. M. (2011). Critical evaluation of folate data in European and international databases: recommendations for standardization in international nutritional studies. *Molecular nutrition & food research, 55*(1), 166-180.

Cavuoto, P., & Fenech, M. F. (2012). A review of methionine dependency and the role of methionine restriction in cancer growth control and life-span extension. *Cancer treatment reviews, 38*(6), 726-736.

Cellarier, E., Durando, X., Vasson, M., Farges, M., Demiden, A., Maurizis, J., . . . Chollet, P. (2003). Methionine dependency and cancer treatment. *Cancer treatment reviews, 29*(6), 489-499.

Chen, P., Li, C., Li, X., Li, J., Chu, R., & Wang, H. (2014). Higher dietary folate intake reduces the breast cancer risk: a systematic review and meta-analysis. *British journal of cancer, 110*(9), 2327-2338. https://doi.org/10.1038/bjc.2014.155.

Deharveng, G., Charrondiere, U. R., Slimani, N., Southgate, D. A., & Riboli, E. (1999). Comparison of nutrients in the food composition tables available in the nine European countries participating in EPIC. European Prospective Investigation into Cancer and Nutrition. *European journal of clinical nutrition, 53*(1), 60-79.

Ducker, G. S., & Rabinowitz, J. D. (2017). One-Carbon Metabolism in Health and Disease. *Cell Metabolism, 25*(1), 27-42. https://doi.org/10.1016/j.cmet.2016.08.009.

European Food information Resource (EuroFIR). What Makes a Food Composition Database (FCDB)? (2020). https://www.eurofir.org/food-information/what-are-fcdbs/ Accessed 1 March 2020.

Feil, R., & Fraga, M. F. (2012). Epigenetics and the environment: emerging patterns and implications. *Nature Reviews Genetics, 13*(2), 97-109. https://doi.org/10.1038/nrg3142.

Friso, S., & Choi, S.-W. (2002). Gene-nutrient interactions and DNA methylation. *The Journal of nutrition, 132*(8), 2382-2387.

Friso, S., Udali, S., De Santis, D., & Choi, S. W. (2017). One-carbon metabolism and epigenetics. *Mol Aspects Med, 54*, 28-36. https://doi.org/10.1016/j.mam.2016.11.007.

Garcia, V., Rona, R. J., & Chinn, S. (2004). Effect of the choice of food composition table on nutrient estimates: a comparison between the British and American (Chilean) tables. *Public health nutrition, 7*(4), 577-583.

Giavarina, D. (2015). Understanding bland altman analysis. *Biochemia medica: Biochemia medica, 25*(2), 141-151.

Greenfield, H., & Southgate, D. A. (2003). *Food composition data: production, management, and use*: Food & Agriculture Org.

Hakala, P., Knuts, L. R., Vuorinen, A., Hammar, N., & Becker, W. (2003). Comparison of nutrient intake data calculated on the basis of two different databases. Results and experiences from a Swedish-Finnish study. *European journal of clinical nutrition, 57*(9), 1035-1044. https://doi.org/10.1038/sj.ejcn.1601639.

Haytowitz, D. B., Pehrsson, P. R., Smith, J., Gebhardt, S. E., Matthews, R. H., Anderson, B. A., & Analysis. (1996). Key foods: setting priorities for nutrient analyses. *9*(4), 331-364.

Jiménez-Chillarón, J. C., Díaz, R., Martínez, D., Pentinat, T., Ramón-Krauel, M., Ribó, S., & Plösch, T. (2012). The role of nutrition on epigenetic modifications and their implications on health. *Biochimie, 94*(11), 2242-2263. https://doi.org/10.1016/j.biochi.2012.06.012.

Julian-Almarcegui, C., Bel-Serrat, S., Kersting, M., Vicente-Rodriguez, G., Nicolas, G., Vyncke, K., . . . Huybrechts, I. (2016). Comparison of different approaches to calculate nutrient intakes based upon 24-h recall data derived from a multicenter study in European adolescents. *European journal of nutrition, 55*(2), 537-545. https://doi.org/10.1007/s00394-015-0870-9.

Koc, H., Mar, M. H., Ranasinghe, A., Swenberg, J. A., & Zeisel, S. H. (2002). Quantitation of choline and its metabolites in tissues and foods by liquid chromatography/electrospray ionization-isotope dilution mass spectrometry. *Analytical chemistry, 74*(18), 4734-4740.

Leclercq, C., Valsta, L. M., & Turrini, A. (2001). Food composition issues--implications for the development of food-based dietary guidelines. *Public health nutrition, 4*(2b), 677-682.

McKay, J., & Mathers, J. (2011). Diet induced epigenetic changes and their implications for health. *Acta Physiologica, 202*(2), 103-118. https://doi.org/10.1111/j.1748-1716.2011.02278.x.

Nazki, F. H., Sameer, A. S., & Ganaie, B. A. (2014). Folate: Metabolism, genes, polymorphisms and the associated diseases. *Gene, 533*(1), 11-20. https://doi.org/10.1016/j.gene.2013.09.063.

Nicolas, G., Witthöft, C. M., Vignat, J., Knaze, V., Huybrechts, I., Roe, M., . . . Slimani, N. (2016). Compilation of a standardised international folate database for EPIC. *Food Chemistry, 193*, 134-140. https://doi.org/10.1016/j.foodchem.2014.11.044.

Obeid, R. (2013). The metabolic burden of methyl donor deficiency with focus on the betaine homocysteine methyltransferase pathway. *Nutrients, 5*(9), 3481-3495. https://doi.org/10.3390/nu5093481.

Riboli, E., Hunt, K., Slimani, N., Ferrari, P., Norat, T., Fahey, M., . . . Vignat, J. (2002). European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public health nutrition, 5*(6b), 1113-1124. https://doi.org/10.1079/PHN2002394.

Riboli, E., & Kaaks, R. (1997). The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. *International journal of epidemiology, 26*(suppl 1), 6-14. https://doi.org/10.1093/ije/26.suppl_1.S6.

Slimani, N., Deharveng, G., Unwin, I., Southgate, D., Vignat, J., Skeie, G., . . . Ireland, J. (2007). The EPIC nutrient database project (ENDB): a first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. *European journal of clinical nutrition, 61*(9), 1037-1056. https://doi.org/10.1038/sj.ejcn.1602679.

Slimani, N., Ferrari, P., Ocke, M., & Welch, A. (2000). Standardization of the 24-hour diet recall calibration method used in the European Prospective Investigation into Cancer and Nutrition (EPIC): general concepts and preliminary results. *European journal of clinical nutrition, 54*(12), 900-917. https://doi.org/10.1038/sj.ejcn.1601107.

Slimani, N., Kaaks, R., Ferrari, P., Casagrande, C., Clavel-Chapelon, F., Lotze, G., . . . Lauria, C. (2002). European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study: rationale, design and population characteristics. *Public health nutrition, 5*(6b), 1125-1145. https://doi.org/10.1079/PHN2002395.

Ulrich, C. M. (2007). Folate and cancer prevention: a closer look at a complex picture. *The American journal of clinical nutrition, 86*(2), 271-273. https://doi.org/10.1093/ajcn/86.2.271.

Vaask, S., Pomerleau, J., Pudule, I., Grinberga, D., Abaravicius, A., Robertson, A., & McKee, M. (2004). Comparison of the Micro-Nutrica Nutritional Analysis program and the Russian Food Composition Database using data from the Baltic Nutrition Surveys. *European journal of clinical nutrition, 58*(4), 573-579. https://doi.org/10.1038/sj.ejcn.1601848.

Wallace, T. C., Blusztajn, J. K., Caudill, M. A., Klatt, K. C., Natker, E., Zeisel, S. H., & Zelman, K. M. (2018). Choline: The Underconsumed and Underappreciated Essential Nutrient. *Nutrition today, 53*(6), 240-253. https://doi.org/10.1097/nt.0000000000000302.

Wu, X. Y., Cheng, J. N., & Lu, L. (2013). Vitamin B12 and Methionine Deficiencies Induce Genome Damage Measured Using the Cytokinesis-Block Micronucleus Cytome Assay in Human B Lymphoblastoid Cell Lines. *Nutrition and Cancer-an International Journal, 65*(6), 866-873. https://doi.org/10.1080/01635581.2013.802000.

Methodological approaches to compile and validate a food composition database for methyl-group carriers in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study.

Heleen Van Puyvelde, Vickà Versele, Marlène De Backer, Corinne Casagrande, Geneviève Nicolas, Joanna L. Clasen, Cristina Julián, Guri Skeie, Maria-Dolores Chirlaque, Yahya Mahamat-Saleh, Pilar Amiano Etxezarreta, Sara Pauwels, Lode Godderis, Marc J. Gunter, Koen Van Herck, Inge Huybrechts, on behalf of the EPIC collaborators

Journal: Food Chemistry

Corresponding author:
Inge Huybrechts, PhD.
International Agency for Research on Cancer (IARC), World Health Organisation
Nutritional Epidemiology Group (NME\NEP)
150 Cours Albert Thomas, 69372 Lyon Cedex 08, France
Email: huybrechtsi@iarc.fr

Appendix 1: List of food composition databases used to compile the methyl-group carrier database

Appendix 1: List of food composition databases used to compile the methyl-group carrier database

Country	Database	Components	Number of food items (%)	Web-address
United States	U.S. FCDB - National Nutrient Database for Standard Reference of the U.S. Department of Agriculture, Release 26 (October 2013 revision)	Total Folate Choline Betaine Methionine	8463 7330 (87%) 4511 (53%) 2005 (24%) 5019 (59%)	https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/nutrient-data-laboratory/docs/sr26-home-page/
Canada	Canadian FCDB - Canadian Nutrient File, 2010	Total Folate Choline Betaine Methionine	5807 5134 (88%) 2415 (42%) 865 (15%) 4039 (70%)	https://food-nutrition.canada.ca/cnf-fce/index-eng.jsp
Germany	German FCDB –Bundes Lebensmittel Schlüssel, version 3.01 (2010)	Total Folate Methionine	10 185 10 185 (100%)[a] 10 185 (100%)[a]	https://www.blsdb.de/
Denmark	Danish FCDB - Danish Food Composition Databank, version 7.01 (March 2009)[b]	Total Folate Methionine	1049 838 (80%) 739 (70%)	http://www.foodcomp.dk/v7/fcdb_aboutfooddata_vitamins.asp

[a] Missing values were replaced by the existing values of the same food group or the same group of constituents

[b] Folate values should be used with caution due to use of an inadequate microbiological assay which systematically provided high folate values (Nicolas et al., 2016)

Methodological approaches to compile and validate a food composition database for methyl-group carriers in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study.

Heleen Van Puyvelde, Vickà Versele, Marlène De Backer, Corinne Casagrande, Geneviève Nicolas, Joanna L. Clasen, Cristina Julián, Guri Skeie, Maria-Dolores Chirlaque, Yahya Mahamat-Saleh, Pilar Amiano Etxezarreta, Sara Pauwels, Lode Godderis, Marc J. Gunter, Koen Van Herck, Inge Huybrechts, on behalf of the EPIC collaborators

Journal: Food Chemistry

Corresponding author:
Inge Huybrechts, PhD.
International Agency for Research on Cancer (IARC), World Health Organisation
Nutritional Epidemiology Group (NME\NEP)
150 Cours Albert Thomas, 69372 Lyon Cedex 08, France
Email: huybrechtsi@iarc.fr

Appendix 2: Paired sample t-test by country for individual dietary folate intake (µg/day) between the methyl-group carrier database (MGDB) and the EPIC nutrient database (ENDB)

Appendix 2: Paired sample t-tests and mean differences by country for individual dietary folate intake (µg/day) between the newly compiled MGDB and the ENDB

		24-HDR				**DQ**			
	Database	**N**	**Mean** (µg/day)	**SD**	**Mean** ∆ (%)*#	**N**	**Mean** (µg/day)	**SD**	**Mean** ∆ (%)*#
Total	MGDB	36 994	325.91	159.30	-60.66	504 247	354.56	127.84	-46.01
	ENDB	36 994	265.25	137.83	(-20.2%)	504 247	308.55	120.14	(-13.88%)
France	MGDB	4735	345.37	149.65	-54.10	73,035	435.43	131.61	-77.88
	ENDB	4735	291.27	142.60	(-17.00%)	73,035	357.55	105.93	(-19.64%)
Italy	MGDB	3961	361.13	193.57	-99.74	45,908	372.38	130.75	-107.07
	ENDB	3961	261.39	150.22	(-32.04%)	45,908	265.31	86.79	(-33.58%)
Spain	MGDB	3220	391.67	209.94	-95.60	40,624	420.59	151.06	-103.71
	ENDB	3220	296.06	160.27	(-27.80%)	40,621	316.88	112.71	(-28.13%)
United Kingdom	MGDB	1315	357.24	154.66	-35.26	81,097	396.17	129.26	12.58
	ENDB	1315	321.98	140.06	(-10.38%)	81,097	408.76	157.68	(3.13%)
The Netherlands	MGDB	4567	326.52	146.32	-54.65	39,037	326.33	86.30	-39.14
	ENDB	4567	271.87	150.64	(-18.27%)	39,037	287.19	74.96	(-12.76%)
Greece	MGDB	2930	297.39	182.30	-24.36	27,476	350.60	111.96	-11.31
	ENDB	2930	273.03	146.83	(-8.54%)	27,476	339.29	109.89	(-3.28%)
Germany	MGDB	4418	330.16	160.34	-82.56	52,013	334.17	96.47	-87.94
	ENDB	4418	247.60	146.03	(-28.58%)	52,013	246.22	70.57	(-30.30%)
Sweden	MGDB	6132	287.67	114.95	-64.66	52,750	285.89	97.84	-41.76
	ENDB	6132	223.01	94.23	(-25.32%)	52,750	244.12	81.41	(-15.76%)
Denmark	MGDB	3918	304.54	118.91	-19.61	55,860	308.62	82.38	-0.72
	ENDB	3918	284.93	113.03	(-6.65%)	55,860	307.90	87.48	(-0.23%)
Norway	MGDB	1798	267.75	117.57	-44.14	36,448	236.07	67.14	-20.58
	ENDB	1798	223.61	92.43	(-17.97%)	36,448	215.49	60.83	(-9.12%)

* Mean difference (%): The MGDB mean minus the ENDB mean (divided by their arithmetic mean [*100%])

[#] Statistical difference $p < 0.001$ for the paired sample t-test

Abbreviations: MGDB: methyl-group carrier database; ENDB: EPIC nutrient database; N: number; SD: standard difference; Mean ∆: mean difference; 24-HDR: 24-hour dietary recall; DQ: dietary questionnaire

Methodological approaches to compile and validate a food composition database for methyl-group carriers in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study.

Heleen Van Puyvelde, Vickà Versele, Marlène De Backer, Corinne Casagrande, Geneviève Nicolas, Joanna L. Clasen, Cristina Julián, Guri Skeie, Maria-Dolores Chirlaque, Yahya Mahamat-Saleh, Pilar Amiano Etxezarreta, Sara Pauwels, Lode Godderis, Marc J. Gunter, Koen Van Herck, Inge Huybrechts, on behalf of the EPIC collaborators

Journal: Food Chemistry

Corresponding author:
Inge Huybrechts, PhD.
International Agency for Research on Cancer (IARC), World Health Organisation
Nutritional Epidemiology Group (NME\NEP)
150 Cours Albert Thomas, 69372 Lyon Cedex 08, France
Email: huybrechtsi@iarc.fr

Appendix 3: Pearson’s correlation coefficients by country for individual dietary folate intake (µg/day) between the methyl-group carrier database (MGDB) and the EPIC nutrient database (ENDB)

Appendix 3: Pearson's correlation coefficients by country for individual dietary folate intake (µg/day) between the MGDB and the ENDB

	Total	France	Italy	Spain	United Kingdom	The Netherlands	Greece	Germany	Sweden	Denmark	Norway
24-HDR	0.73	0.70	0.71	0.70	0.69	0.77	0.84	0.71	0.73	0.80	0.80
DQ	0.81	0.79	0.80	0.77	0.88	0.85	0.98	0.89	0.82	0.90	0.96

All correlations are statistically significant ($p < 0.001$)

Abbreviations: MGDB: methyl-group carrier database; ENDB: EPIC nutrient database; 24-HDR: 24-hour dietary recall; DQ: dietary questionnaire

Methodological approaches to compile and validate a food composition database for methyl-group carriers in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study.

Heleen Van Puyvelde, Vickà Versele, Marlène De Backer, Corinne Casagrande, Geneviève Nicolas, Joanna L. Clasen, Cristina Julián, Guri Skeie, Maria-Dolores Chirlaque, Yahya Mahamat-Saleh, Pilar Amiano Etxezarreta, Sara Pauwels, Lode Godderis, Marc J. Gunter, Koen Van Herck, Inge Huybrechts, on behalf of the EPIC collaborators

Journal: Food Chemistry

Corresponding author:
Inge Huybrechts, PhD.
International Agency for Research on Cancer (IARC), World Health Organisation
Nutritional Epidemiology Group (NME\NEP)
150 Cours Albert Thomas, 69372 Lyon Cedex 08, France
Email: huybrechtsi@iarc.fr

Appendix 4: Weighted Kappas by country for individual dietary folate intake (µg/day) between the methyl-group carrier database (MGDB) and the EPIC nutrient database (ENDB)

Appendix 4: Weighted Kappas by country for individual dietary folate intake (µg/day) between the MGDB and the ENDB

	24-HDR						DQ					
	Classified into the same Q (%)	Classified into the adjacent Q (%)	Weighted κ	SE	CI lower	CI upper	Classified into the same Q (%)	Classified into the adjacent Q (%)	Weighted κ	SE	CI lower	CI upper
Total	46.17	39.83	0.56	0.003	0.55	0.56	50.59	40.53	0.63	0.001	0.63	0.63
France	44.03	40.53	0.53	0.007	0.51	0.54	47.68	40.48	0.59	0.002	0.59	0.59
Italy	44.18	39.96	0.53	0.008	0.52	0.55	47.85	41.16	0.60	0.002	0.60	0.61
Spain	47.92	39.13	0.58	0.009	0.56	0.59	50.24	40.20	0.62	0.002	0.62	0.63
United Kingdom	44.87	38.18	0.52	0.014	0.50	0.55	58.07	35.76	0.70	0.002	0.69	0.70
The Netherlands	47.03	39.37	0.57	0.007	0.55	0.58	56.52	36.63	0.68	0.003	0.67	0.68
Greece	56.76	36.35	0.68	0.009	0.66	0.70	77.08	22.73	0.86	0.003	0.85	0.86
Germany	46.11	39.81	0.55	0.008	0.54	0.56	57.73	36.77	0.70	0.002	0.69	0.70
Sweden	45.82	39.78	0.55	0.006	0.54	0.56	54.20	37.96	0.66	0.002	0.65	0.66
Denmark	50.79	38.74	0.62	0.008	0.60	0.63	59.83	35.70	0.72	0.002	0.72	0.72
Norway	52.06	39.66	0.64	0.012	0.61	0.66	47.68	40.48	0.81	0.003	0.80	0.81

Abbreviations: MGDB: methyl-group carrier database; ENDB: EPIC nutrient database; Q: quintile; κ: kappa; SE: standard error; CI: 95% confidence interval