Nutrition, Metabolism & Cardiovascular Diseases (2022) **32**, 80–89

Available online at www.sciencedirect.com

Nutrition, Metabolism & Cardiovascular Diseases

journal homepage: www.elsevier.com/locate/nmcd

Evaluation of protein and amino acid intake estimates from the EPIC dietary questionnaires and 24-h dietary recalls using different food composition databases

Isabel Iguacel [a,b,c,d,e,*], Aurora Perez-Cornago [f], Julie A. Schmidt [f], Heleen Van Puyvelde [a,g], Ruth Travis [f], Corinne Casagrande [a], Genevieve Nicolas [a], Elio Riboli [h], Elisabete Weiderpass [a], Eva Ardanaz [i,j,k], Aurelio Barricarte [i,j,k], Stina Bodén [l], Eleonora Bruno [m], Ana Ching-López [n,o], Dagfinn Aune [h,p,q], Torill E. Jensen [r], Ulrika Ericson [s], Ingergerd Johansson [t], José Ma Huerta [u,v], Verena Katzke [w], Tilman Kühn [w], Carlotta Sacerdote [x], Matthias B. Schulze [y,z], Guri Skeie [r], Stina Ramne [s], Heather Ward [h], Marc J. Gunter [a], Inge Huybrechts [a]

[a] *International Agency for Research on Cancer, Nutrition and Metabolism Branch, 69372 Lyon Cedex 08, France*
[b] *Department of Physiatry and Nursing, Faculty of Health Sciences, University of Zaragoza, Zaragoza, Spain*
[c] *Instituto Agroalimentario de Aragón, Zaragoza, Spain*
[d] *Instituto de Investigación Sanitaria Aragón, Zaragoza, Spain*
[e] *Centro de Investigación Biomédica en Red de Fisiopatología de la Obesidad y Nutrición, Zaragoza, Spain*
[f] *Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom*
[g] *Department of Public Health and Primary Care, Faculty of Medicine and Health Sciences, Ghent University, 9000 Ghent, Belgium*
[h] *Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom*
[i] *Navarra Public Health Institute, Pamplona, Spain*
[j] *Navarra Institute for Health Research (IdiSNA), Pamplona, Spain*
[k] *CIBER Epidemiology and Public Health (CIBERESP), Madrid, Spain*
[l] *Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden*
[m] *Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Via Venezian, 1, 20133 Milano, Italy*
[n] *Escuela Andaluza de Salud Pública (EASP), Granada, Spain*
[o] *Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain*
[p] *Department of Nutrition, Bjørknes University College, Oslo, Norway*
[q] *Department of Endocrinology, Morbid Obesity and Preventive Medicine, Oslo University Hospital, Oslo, Norway*
[r] *Department of Community Medicine, Faculty of Health Sciences, UiT-The Arctic University of Norway, Tromsø, Norway*
[s] *Department of Clinical Sciences Malmö, Medical Faculty, Lund University, Sweden*
[t] *Department of Odontology, Section of Cardiology, Biobank Research, Public Health and Clinical Medicine, Umeå University, Umeå, Sweden*
[u] *Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain*
[v] *CIBER Epidemiología y Salud Pública (CIBERESP), Spain*
[w] *German Cancer Research Center (DKFZ), Foundation under Public Law, Heidelberg, Germany*
[x] *Unit of Cancer Epidemiology, Città della Salute e della Scienza University-Hospital and Center for Cancer Prevention (CPO), Turin, Italy*
[y] *Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany*
[z] *Institute of Nutritional Sciences, University of Potsdam, Nuthetal, Germany*

Received 12 April 2021; received in revised form 21 July 2021; accepted 9 September 2021
Handling Editor: A. Naska
Available online 20 September 2021

KEYWORDS
Amino acid intakes;
Dietary questionnaire;
Food composition

Abstract *Background and aims:* This study aimed to expand the European Prospective Investigation into Cancer and Nutrition (EPIC) nutrient database (ENDB) by adding amino acid (AA) values, using the U.S. nutrient database (USNDB). Additionally, we aimed to evaluate these new protein and AA intake estimates from the EPIC dietary questionnaires (DQ) and 24-h dietary recalls (24-HDR) using different matching procedures.

Abbreviations: AA, amino acids; EPIC, European Prospective Investigation into Cancer and Nutrition; USNDB, United States nutrient database; DQ, dietary questionnaire; 24-HDR, 24-h dietary recall; ENDB, EPIC nutrient database; DR, Dietary record; FFQ, Food Frequency questionnaire.

* Corresponding author. International Agency for Research on Cancer, Nutrition and Metabolism Branch, 69372 Lyon Cedex 08, France.
E-mail address: iguacel@unizar.es (I. Iguacel).

https://doi.org/10.1016/j.numecd.2021.09.012

tables;
24-h dietary recall;
Validity

Methods and results: Dietary energy, protein and AA intakes were assessed via DQ and 24-HDR by matching with the USNDB food composition table. Energy and protein intakes calculated using USNDB matching were compared with those calculated using ENDB, that uses country specific food composition tables. Pearson correlations, Cohen's weighted kappa statistic and Bland −Altman plots were used to compare data resulting from USNDB matching with our reference from ENDB matching.

Very high correlations were found when comparing daily energy (r = 0.99) and dietary protein intakes (r = 0.97) assessed via USNDB with those obtained via ENDB (matching for DQ and 24-HDR). Significant positive correlations were also found with energy and protein intakes acquired via 24-HDRs in the EPIC calibration sample.

Conclusion: Very high correlations between total energy and protein intake obtained via the USDA matching and those available in ENDB suggest accuracy in the food matching. Individual AA have been included in the extended EPIC Nutrient database that will allow important analyses on AA disease prospective associations in the EPIC study.

Introduction

The potential role of several amino acids (AA) in the development of various diseases including renal failure, liver cirrhosis, diabetes, and cancer have been elucidated in recent studies [1,4,6,8,22]. However, mechanistic pathways and effects of dietary AA intakes on blood levels and disease outcomes remain unclear.

An important limitation when investigating associations between AA intakes and health outcomes and the potential underlying mechanisms, is the lack of detailed data on individual dietary AA intakes in large-scale cohorts, since National food composition tables rarely include data on AA composition. One of the exceptions is the U.S. nutrient database (USNDB); i.e. the National Nutrient Database for Standard Reference of the United States developed at the United States Department of Agriculture (USDA). The USDNB includes a large number of food and recipe items from various countries and eating cultures (>8000 food items in the USNDB release 26 [October 2013]) and uses standard reference analytical methods to obtain the nutritional values [20].

Therefore, we used the USNDB to estimate AA intakes among participants of the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort in order to allow future in-depth analysis on the roles of AA intakes and their impact on health outcomes. However, before any evaluation of estimated intakes in relation to disease outcomes can be done, the validity of this food matching from dietary questionnaires (DQ) and 24-h dietary recalls (24-HDR) needed to be assessed to be able to account for measurement error in cohort studies with self-reported dietary intake.

Although AAs had not been included in the ENDB, the EPIC-Oxford cohort had already added AA data to their local dietary intake database. This process was conducted by a local researcher in Oxford [13] allowing an inter-rater reliability assessment for the AA intake data retrieved for EPIC-Oxford when matching the USDA data with the food list of all EPIC centers.

Consequently, the aims of the current study were (1) to extend the EPIC Nutrient Database (ENDB) with AA estimates among participants of the EPIC cohort using the USDA database and (2) to compare estimated intakes of energy and protein included in the ENDB and USDA database as quality control. Furthermore, we aim (3) to calculate the relative validities of AA intakes derived from USDA nutrient data added to the DQ and 24-HDR food intake estimates. In addition, (4) an inter-rater reliability assessment was performed for the AA matching with the EPIC-Oxford cohort dietary data.

Methods

Study design and participants

EPIC is a large on-going multicentre prospective cohort study consisting of 521,324 adults (366,521 women and 153,437 men) aged 25−70 years from whom diet, and lifestyle data were collected at baseline. The objective of this cohort is to investigate the role of diet, lifestyle, metabolic factors and genetics in cancer development as well as other chronic diseases in a European sample. The participants were enrolled between 1992 and 2000 from 23 centres in 10 European countries: Denmark, France, Germany, Greece, Italy, The Netherlands, Norway, Spain, Sweden and the United Kingdom [11]. The rationale, study population and data collection have been described elsewhere [12]. All participants provided written informed consent and the ethical review boards from the International Agency for Research on Cancer (IARC) and from all local centres approved the study. Participants with missing information on dietary intake were excluded from the analysis. The final study population for the dietary

assessment within the EPIC study included 504,245 participants (70.8% females) after excluding participants with missing information on more than 80% of the relevant questions of the DQ (n = 6837) and with implausible energy intakes (individuals in the highest and lowest 1% of the distribution for the ratio of energy intake to estimated energy requirement; n = 10,242) (Supplementary Fig. 1).

Assessment of diet

Habitual dietary intake data were assessed at recruitment using a centre-specific quantitative DQ or a semi-quantitative FFQ developed and validated in each country/centre. These were completed by 514,487 participants. Following a highly standardized procedure, a computer-assisted, 24-HDR interview program (EPIC-soft) was used to conduct a single interactive, face-to-face (or telephone in Norway) dietary interview [18]. The 24-HDRs were collected from a representative sample (calibration cohort n = 36,994) from the entire EPIC cohort [19] and from which 36,978 subjects had valid information (both, a DQ and a 24-HDR). Nutrient values from the national food composition databases of the 10 EPIC countries were used to compile the ENDB [16]. In particular, a total of 550−1500 foods derived from about 37,000 standardized EPIC 24-HDRs were calculated using foods available in the 10 national Nutrient Databases (NDBs). The extra foods that were not included in the 24-HDR food list but generated from the DQs were then matched to extra NDB items or treated as generic or recipes. In the absence of a standardized European nutrient database, the average nitrogen (N) and energy intakes were calculated using country-specific food composition tables for both dietary assessment methods [17]. More detailed information about this compiling process can be found elsewhere [17].

Dietary values for AAs were not included in the ENDB, therefore data from the USNDB was used to estimate individual dietary intakes of AAs for the EPIC cohort. Adding USNDB data to the EPIC dietary intake data was done following the same standardized procedures as described for the ENDB project [17]. Here, 19 individual AAs (Tables 2a and 2b) were included in the EPIC nutrient database by adding the USNDB food composition data (SR 26−28) [21] following standardized procedures as used in the ENDB approach. In brief, the USNDB food composition data were first added to the food list derived from the 24-HDR, which was then used as a basis for the matching with the foods reported in the DQ. Specific foods and recipes that were not included in the USNDB were decomposed into ingredients available in the USNDB table. Various quality controls (e.g. double data entry/matching by two independent dietitians; checking of outliers within food groups, etc.) were carried out to optimize the quality of the food matching and to avoid errors.

Validity intakes and amino acid intakes

Total energy and protein intakes that had already been assessed and validated in the frame of the ENDB project [11] were compared with the energy and protein intakes assessed using the USNDB to evaluate the potential bias due to the matching and calculations procedures with different food composition tables. The AA intakes for both DQ data and 24-HDR data estimated by the USNDB were compared with each other and with AA intake estimates obtained previously for 57,397 EPIC-Oxford participants. The latter AA intakes were obtained using independent matching of DQ food items to the same USNDB [13,21].

Statistical analysis

Demographic characteristics of the study participants were reported as mean and standard deviation (SD) for continuous variables and the percentages for categorical variables. Additionally, the mean, SD, and percentiles 25th, 50th and 75th of daily intake for dietary AA calculated via USNDB as well as the energy (kcal) and total proteins (g) estimated by both methods (ENDB and USNDB) were calculated.

For the relative validation analysis specified below, all the AA intakes were logarithmically transformed to better approximate the normal distribution.

Energy and total protein intakes obtained via the USNDB were compared with the energy and total protein intakes estimated by the ENDB (validated reference values) using three statistical methods: Pearson correlations, Cohen's weighted kappa statistic [7] and Bland−Altman plots with 95% limits of agreement [3]. The comparative analyses were performed for both, the DQ data (using the full EPIC cohort, N = 504,245) as well as the 24-HDR data (using the calibration subsample, N = 36,994).

In addition, Pearson's correlations were used to compare the AA intakes estimated through the USNDB between DQ data and 24-HDR data.

Because AA were not included in the ENDB, the nutrient database from EPIC-Oxford was used to allow for comparison of the AA estimates with the USNDB. The estimation of AA intakes for EPIC-Oxford participants was performed independently from the current study by a local researcher in Oxford also using USNDB data; the procedures used were published before by Schmidt et al. [13]. The dietary intakes of AA calculated only for the EPIC-Oxford cohort were compared with those of the main analysis using Pearson's correlations for the 57,397 EPIC-Oxford participants.

Conventional two-sided P-values are shown, but all results have been interpreted after allowance for multiple testing using the Bonferroni method; the per-test significance level was 0.05 divided by the number of tests (p = 0.05/20 = 0.0025). All analyses were conducted using SAS version 9.4.

Results

Table 1 shows the main demographic characteristics (age, Body Mass Index (BMI), sex and education) of the study participants in the full EPIC cohort (DQs; N = 504,245) and in the EPIC calibration subcohort (24-HDRs; N = 36,994).

Table 1 Demographic characteristics of the study participants included in the full EPIC cohort (with Diet Questionnaires – DQs) and in the EPIC calibration subcohort (24-h Dietary Recall – 24-HDRs).

Demographic characteristics	N = 504,245 (with DQ)		N = 36,994 (24-h recalls)	
	Mean	SD	Mean	SD
Age (years)	51.41	9.96	53.99	8.76
BMI (kg/m^2)	25.40	4.27	25.81	4.25
	n	%	n	%
Sex				
Male	147,259	29.2	13,486	36.45
Female	356,986	70.8	23,508	63.55
Education				
Low	150,324	29.81	12,366	33.43
Medium	215,136	42.66	15,739	42.54
High	120,157	23.82	8363	22.61
Not reported or missing	18,628	3.69	526	1.42

Abbreviations: BMI, Body Mass Index; DQs, Diet Questionnaires; 24-HDRs, 24-Hour Dietary Recall; SD, Standard Deviation.

Among the 504,245 participants with DQ information, most were females (70.8%) with a medium level of education (42.6%) and average BMI of 25.4 kg/m^2.

Tables 2a and 2b present mean (total and stratified by sex), SD, 25th, 50th and 75th percentile of estimated dietary AA by the USNDB from DQ (Table 2a) and 24-HDR (Table 2b). Overall, dietary AA from DQ data ranged from 0.07 g (hydroxy-proline) to 13.56 g (glutamic acid) and from 24-HDR data ranged from 0.07 g (hydroxy-proline) to 13.22 g (glutamic acid). Differences in AA intakes measured by DQ and 24-HDR were very small (<10% of mean intake for both, mean and median intake estimates),

Table 2a Mean, SD, 25th, 50th and 75th percentile of dietary amino acid (AA) intakes (USDA) of EPIC participants with dietary questionnaire (DQ) information (N = 504,245; 147,259 men & 356,986 Females).

Dietary AA intakes (in g)	Mean (male)	Mean (female)	Total mean	SD	P25	P50	P75
Alanine	3.65	3.05	3.23	1.13	2.43	3.09	3.87
Arginine	4.08	3.43	3.62	1.23	2.75	3.46	4.31
Aspartic acid	7.09	6.08	6.37	2.05	4.93	6.14	7.54
Cystine	0.97	0.79	0.84	0.29	0.64	0.81	1.00
Glutamic acid	15.45	12.79	13.56	4.57	10.38	12.93	16.03
Glycine	3.10	2.47	2.65	0.96	1.98	2.53	3.18
Histidine	2.25	1.90	2.00	0.70	1.51	1.92	2.40
Hydroxyproline	0.09	0.06	0.07	0.05	0.03	0.06	0.09
Isoleucine	3.52	3.01	3.16	1.09	2.40	3.02	3.76
Leucine	6.20	5.32	5.58	1.92	4.24	5.34	6.64
Lysine	5.51	4.79	5.00	1.81	3.73	4.80	6.03
Methionine	1.78	1.54	1.61	0.57	1.21	1.54	1.93
Phenylalanine	3.54	3.01	3.16	1.06	2.42	3.03	3.75
Proline	5.52	4.68	4.92	1.80	3.67	4.65	5.86
Serine	3.71	3.18	3.34	1.12	2.55	3.19	3.96
Threonine	3.08	2.63	2.76	0.94	2.10	2.65	3.29
Tryptophan	0.92	0.78	0.82	0.28	0.62	0.78	0.97
Tyrosine	2.81	2.45	2.56	0.91	1.93	2.44	3.05
Valine	4.20	3.63	3.79	1.29	2.90	3.64	4.51

Table 2b Mean, SD, 25th, 50th and 75th percentile of dietary AA intakes (USDA) of EPIC participants with 24-h dietary recall (24-HDR) information (N = 36,994; 13,486 men & 23,508 Females).

Dietary AA intakes (in g)	Mean (male)	Mean (female)	Mean	SD	P25	P50	P75
Alanine	3.67	2.76	3.09	1.5	2.1	2.86	3.84
Arginine	4.09	3.10	3.46	1.7	2.3	3.18	4.29
Aspartic acid	6.96	5.41	5.97	2.7	4.1	5.57	7.35
Cysteine	1.00	0.74	0.83	0.4	0.6	0.77	1.03
Glutamic acid	15.54	11.89	13.22	5.7	9.3	12.4	16.3
Glycine	3.13	2.27	2.58	1.3	1.7	2.36	3.23
Histidine	2.25	1.74	1.93	0.9	1.3	1.8	2.39
Hydroxyproline	0.09	0.06	0.07	0.1	0	0.01	0.09
Isoleucine	3.50	2.71	3	1.4	2.1	2.79	3.7
Leucine	6.17	4.82	5.32	2.4	3.7	4.98	6.55
Lysine	5.50	4.33	4.75	2.3	3.2	4.4	5.94
Methionine	1.78	1.39	1.53	0.7	1	1.43	1.91
Phenylalanine	3.51	2.73	3.02	1.3	2.1	2.83	3.71
Proline	5.47	4.31	4.74	2.1	3.3	4.45	5.86
Serine	3.70	2.89	3.19	1.4	2.3	3	3.92
Threonine	3.06	2.36	2.62	1.2	1.8	2.44	3.23
Tryptophan	0.91	0.70	0.78	0.4	0.5	0.73	0.96
Tyrosine	2.78	2.22	2.42	1.1	1.7	2.28	3.01
Valine	4.15	3.26	3.59	1.6	2.5	3.37	4.42

Abbreviations: AA, Amino acids; EPIC, European Prospective Investigation into Cancer and Nutrition; DQs, Diet Questionnaires; 24-HDRs, 24-Hour Dietary Recall; SD, Standard Deviation.

although standard deviations were slightly higher in the 24-HDRs compared with the DQs.

Table 3 shows mean, 25th, 50th and 75th percentile, mean difference, Pearson correlations and weighted Kappas of daily energy and total protein intakes calculated using both ENDB and USNDB. For the DQ there was a lower mean daily energy intake (2070.16 kcal/d) but a higher protein intake (86.47 g/d) using the ENDB compared with the USNDB (2129.32 kcal/d and 81.94 g/d) respectively. That is, there was only a minor mean difference in energy (−59.16 kcal/d) and protein (4.53 g/d) intake values obtained via the USNDB and those generated by the ENDB [17] (validated reference method) for the DQs. Very high correlations were also found between the two methods for both energy (0.99) and total protein (0.97). In addition, there was a very good agreement between the protein (weighted Kappa = 0.84) and energy (weighted Kappa = 0.89) intakes obtained via USDNB and those generated in the ENDB (Table 3) [7]. Moreover, Bland–Altman plots of energy and protein intakes also showed narrow limits of agreement (−287, 169 kcal/d and −9, 18 g/d for energy and protein intakes respectively) (Fig. 1).

Results obtained for the 24-HDRs were very similar to those reported for the DQs (Table 3). The 24-HDR showed very high correlations but slightly lower than the DQs when comparing the energy (r = 0.96) and protein (r = 0.93) intakes estimated by the ENDB and USNDB. Also, a very good agreement between the protein (weighted Kappa = 0.77) and energy (weighted Kappa = 0.83) intakes obtained via ENDB and USNDB was observed (Table 3) [7]. Larger limits of agreement were obtained for energy (−483, 439 kcal/d) and protein (−20, 32 g/d) intakes

Table 3 Mean, 25th, 50th and 75th percentile, mean difference, Pearson correlations and weighted Kappas of daily energy and total protein intakes using the standardized EPIC Nutrient Database (ENDB) versus the U.S. Nutrient Database (USNDB): EPIC participants (N = 504,245 with DQs information and N = 36,994 with 24-HDR information).

DQs	ENDB					USNDB					Mean difference	Limits of agreement	Pearson correlations	Weighted Kappa
	Mean	SD	P25	P50	P75	Mean	SD	P25	P50	P75				
Energy (kcal)	2070.16	617.82	1627.30	1992.20	2430.06	2129.32	623.84	1683.80	2053.93	2492.57	−59.16	(−287; 169)	0.99	0.89
Total proteins (g)	86.47	27.34	67.18	83.38	102.17	81.94	25.47	64.08	79.01	96.35	4.53	(−9; 18)	0.97	0.84
24-HDR	ENDB					USNDB					Mean difference	Limits of agreement	Pearson correlations	Weighted Kappa
	Mean	SD	P25	P50	P75	Mean	SD	P25	P50	P75				
Energy (kcal)	2093.33	792.24	1542.97	1979.74	2513.93	2114.95	794.75	1562.62	2002.81	2532.93	−21.62	(−483; 439)	0.96	0.83
Total proteins (g)	83.48	34.95	59.83	78.13	100.63	77.47	31.27	56.03	72.94	93.59	6.01	(−20; 32)	0.93	0.77

Abbreviations: ENDB, EPIC Nutrient Database; USNDB, U.S. nutrient database; EPIC, European Prospective Investigation into Cancer and Nutrition; DQs, Diet Questionnaires; 24-HDRs, 24-Hour Dietary Recall; P25, P50 and P75, Percentile 25, 50 and 75; SD, Standard Deviation.

in the Bland–Altman plots derived from the 24-HDR compared with those derived from the DQ (Fig. 2).

In Figs. 1 and 2, the Bland–Altman plots for the comparison of energy and protein obtained in the ENDB versus USNDB using the DQ data are satisfactory although there appears to be a tendency of over-dispersion (non-constant variance) for higher intakes (over 4000 kcal or 200 g for energy and protein respectively).

This is even more apparent in the Bland–Altman plots for the 24-HDR data thus suggesting the need for log transformation. It appears also that in the 24-HDR there is an underestimation from the ENDB compared with USDA for energy intake estimates.

Significant and positive correlations ($r = 0.34–0.42$; $p < 0.0001$) were found for the different AA intakes between the DQ and the single 24-HDR from the EPIC calibration study (Table 4). In particular, proline and glutamic acid ($r = 0.40$; $p < 0.0001$) as well as cysteine ($r = 0.38$; $p < 0.0001$) had the highest correlations while arginine and lysine ($r = 0.34$; $p < 0.0001$) had the lowest. Finally, the correlations between the DQ and the single 24-HDR were moderate for total energy ($r = 0.42$) and total protein intake ($r = 0.38$).

Lastly, from DQ data, very high correlations (r for all AA ≥ 0.90; $p < 0.0001$) were obtained between AA intakes assessed by the USNDB and those estimated independently for EPIC-Oxford by Schmidt et al. based on 57,397 EPIC-Oxford participants (Table 4) [13], demonstrating strong inter-investigator reliability.

Discussion

The objective of this paper is to compare the estimated intakes of energy, protein and AAs using two different food composition databases, i.e. the ENDB vs. USNDB and the EPIC-Oxford NDB vs. USNDB. To date, most of the national food composition databases of the EPIC countries did not contain nutritional values for AA. As the USNDB includes foods and recipe items from various countries and eating cultures [20] the matching process was feasible between the USNDB with the EPIC food consumption data. The relative validation analyses were performed for energy and protein intakes (as AA were not available in the ENDB), using three different statistical methods (Pearson correlations, Cohen's weighted kappa statistic and Bland–Altman plots with 95% limits of agreement) to compare the protein and energy intake data estimated by the ENDB and the USNDB [17]. Results have shown good relative validity of the food matching performed with the USNDB food composition data. Because AA were not included in the ENDB, the nutrient database from EPIC-Oxford was used to allow for comparison of the AA estimates with the USNDB, showing very high correlations between the two independent estimations of AA intakes. As such, these quality controls and relative validity analyses of the protein, energy and AA intake data obtained through the EPIC DQ, suggested good quality of the protein and AA intake data calculated for the full EPIC cohort.

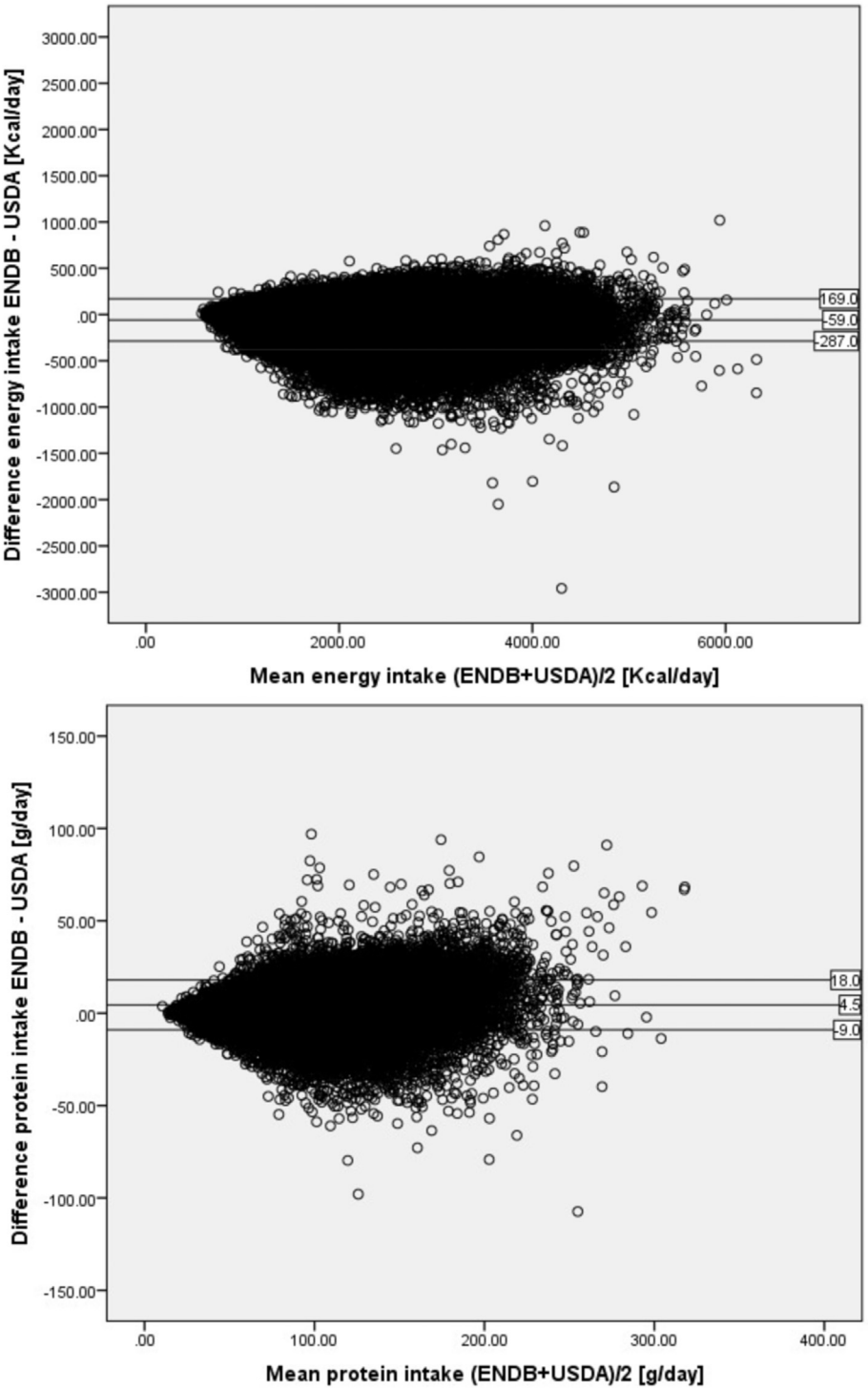

Figure 1 Bland and Altman plots based upon the DQ data, representing the mean difference and limits of agreement for energy and protein intake between the EPIC nutrient database (ENDB) and the U.S. nutrient database (USNDB) for the full EPIC cohort (N = 504,245). Abbreviations: ENDB, EPIC Nutrient Database; EPIC, European Prospective Investigation into Cancer and Nutrition; DQs, Diet Questionnaires; USNDB, U.S. nutrient database.

In Japan, Ishihara J and colleagues conducted a validation study in which they demonstrated that validity in estimating AA intakes was low to moderate for the DQ when 28-day weighed dietary records (DR) were used as a reference method [5]. Spearman's rank correlation coefficients between AA intakes from the DQ and DR ranged from 0.15 to 0.52 and the median correlation coefficients were 0.33 for men (n = 102) and 0.25 for women (n = 113) in the internal population, and 0.40 for men (n = 174) and 0.30 (n = 176) for women in the

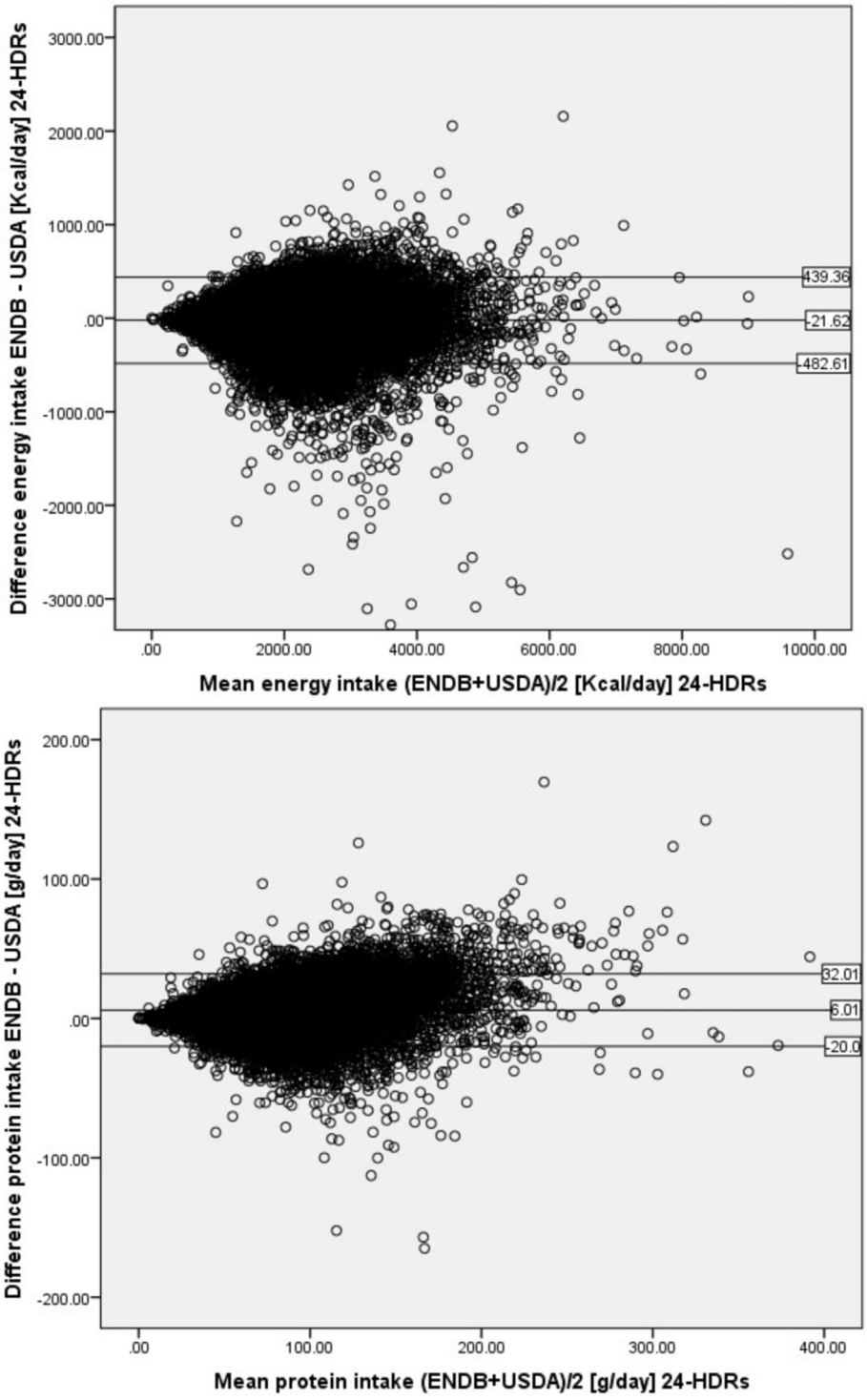

Figure 2 Bland and Altman plots based upon the 24-HDR data, representing the mean difference and limits of agreement for energy and protein intake between the EPIC nutrient database (ENDB) and the U.S. nutrient database (USNDB) for the full EPIC cohort (N = 36,994). Abbreviations: ENDB, EPIC Nutrient Database; EPIC, European Prospective Investigation into Cancer and Nutrition; 24-HDRs, 24-Hour Dietary Recall; USNDB, U.S. nutrient database.

external population (a separate population used to confirm external validity). The authors also concluded that protein was underestimated by the DQ, particularly among men [5].

Analogous to other studies, our mean values of daily energy and total protein intakes estimated from the DQ were slightly higher than those estimated from the 24-HDR and the limits of agreement were better for the DQ

Table 4 Pearson's correlation coefficients between dietary intakes of amino acid (AA) calculated by the U.S. nutrient database (USDA) and for the nutrient database of the EPIC-Oxford study cohort by Schmidt et al. [13], reported for the 24-h dietary recall data (24-HDR) and the dietary questionnaire data (DQ).[a]

	Pearson correlations EPIC-Europe and EPIC-Oxford DQ AA intakes (n = 57,397)	p[b]	Pearson correlations DQ-24-HDR collected from a representative sample from the entire EPIC cohort (n = 36,978)	p[b]
Alanine (g)	0.952	**<0.0001**	0.358	**<0.0001**
Arginine (g)	0.952	**<0.0001**	0.341	**<0.0001**
Aspartic acid (g)	0.959	**<0.0001**	0.358	**<0.0001**
Cysteine (g)	0.897	**<0.0001**	0.384	**<0.0001**
Glutamic acid (g)	0.910	**<0.0001**	0.398	**<0.0001**
Glycine (g)	0.939	**<0.0001**	0.373	**<0.0001**
Histidine (g)	0.952	**<0.0001**	0.353	**<0.0001**
Isoleucine (g)	0.949	**<0.0001**	0.368	**<0.0001**
Leucine (g)	0.949	**<0.0001**	0.361	**<0.0001**
Lysine (g)	0.962	**<0.0001**	0.343	**<0.0001**
Methionine (g)	0.960	**<0.0001**	0.347	**<0.0001**
Phenylalanine (g)	0.939	**<0.0001**	0.375	**<0.0001**
Proline (g)	0.914	**<0.0001**	0.397	**<0.0001**
Serine (g)	0.938	**<0.0001**	0.362	**<0.0001**
Threonine (g)	0.950	**<0.0001**	0.358	**<0.0001**
Tryptophan (g)	0.914	**<0.0001**	0.359	**<0.0001**
Tyrosine (g)	0.948	**<0.0001**	0.351	**<0.0001**
Valine (g)	0.949	**<0.0001**	0.361	**<0.0001**
Energy (kcal)	0.976	**<0.0001**	0.421	**<0.0001**
Total proteins (g)	0.956	**<0.0001**	0.377	**<0.0001**

Abbreviations: AA, Amino acids; EPIC, European Prospective Investigation into Cancer and Nutrition; DQs, Diet Questionnaires; 24-HDRs, 24-Hour Dietary Recall.

[a] All results have been interpreted after allowance for multiple testing using the Bonferroni method; the per-test significance level was 0.05 divided by the number of tests (20).

[b] Conventional P-values are shown, and those marked with in bold were significant after Bonferroni correction (P = 0.0025).

than for the 24-HDR (most likely due to the day-to-day variation in the 24-HDR data) [9]. 24-HDR offers more detailed information than DQs [15] while it does not reflect long term dietary intake, but very similar results were obtained when comparing the estimates between the USNDB and the ENDB.

Even though our results showed a good relative validity of the food matching, when comparing energy and protein from DQ and 24-HDR in Bland–Atman plots, bias seemed to be higher in larger intakes (i.e. over 4000 kcal or 200 g for energy and protein).

Finally, despite the development of a calibration approach to adjust for possible systematic over- or underestimation in dietary intake measurements [19] there seems to be a small underestimation in energy obtained from 24-HDR when using the ENDB compared with USDA.

This study has some limitations. Firstly, matching the European foods with only one database of AA intakes (the USNDB) may have some disadvantages (e.g. lack of traditional local dishes and foods), although the USNDB does include many foods from numerous regions in the world. It should be noted that linking different national (country-specific) food composition tables, as was done in the ENDB, may result in errors and discrepancies in the nutrient content of the Food Composition Databases due to the diverse assays used in nutrient estimation and sampling procedures in the different country-specific food composition tables used [14]. Nevertheless, our study results demonstrated very good comparison between dietary intakes of energy and protein estimated by the USNDB and ENDB. Moreover, one single 24-HDR is not the gold standard for assessing usual dietary intake because of day-to-day variability in human diets. These limitations could result in measurement errors when assessing dietary AA intakes. Despite these limitations, our results have shown positive moderate correlations between the DQ and the single 24-HDR from the EPIC calibration study and an almost perfect correlation for the total energy and protein intakes calculated via the validated reference method (ENDB) versus the USNDB [10]. Indeed, the reference protein and energy intakes obtained from the local/national food composition tables that were included in the ENDB had been validated in a previous study that compared mean protein (= N*6.25) intake and indirectly total energy intake estimated from two methods (24-HDRs and DQ) in comparison with 24-h urinary N [16]. The 24-HDRs and DQ provided good agreement of centre mean total N or energy intakes when compared with urinary N. However, an overall better quantitative agreement existed with urinary N when using a highly standardized 24-HDR than with DQ, probably due to the fact that urine collection is linked to the 24-HDR period [2,16].

There were several strengths to this study. To the best of our knowledge, this is the largest study existing in the

literature comparing the intake of dietary AA using different food composition databases. Moreover, great care was taken when matching the USNDB with the EPIC food list including varied quality control procedures.

In conclusion, individual AAs have now been included in the EPIC nutrient database through matching with the USNDB following standardized procedures. Statistical analysis indicated a good comparability of protein and energy estimation based on the food matching performed with the USNDB compared with our previous ENDB estimation. Our analyses also demonstrated the feasibility of assessing AA intakes from detailed DQ when using standardized and in-depth quality control procedures. The estimated dietary intakes of AA can be used to test possible associations with diseases such as cancer and may in turn help develop our understanding of the role of dietary AA intakes in disease development.

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.

Authorship

The authors' contributions were as follows: I.I and I.H designed the research; I.I, I.H conducted the research; I.I and I.H analysed the data and I.I and I.H, A.C and J.A.S wrote the paper. The manuscript was drafted and prepared, reviewed and revised by all authors. All authors made substantial contributions to the paper and approved the final version of the manuscript.

Availability of data and materials

For information on how to submit an application for gaining access to EPIC data and/or biospecimens, please follow the instructions at http://epic.iarc.fr/access/index.php.

Financial support

The coordination of EPIC is financially supported by International Agency for Research on Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London which has additional infrastructure support provided by the NIHR Imperial Biomedical Research Centre (BRC).

The national cohorts are supported by: Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), Federal Ministry of Education and Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy, Compagnia di SanPaolo and National Research Council (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), Statistics Netherlands (The Netherlands); Health Research Fund (FIS) – Instituto de Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, and the Catalan Institute of Oncology - ICO (Spain); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford) (United Kingdom).

Declaration of competing interest

No authors report conflicts of interest.

Acknowledgments

We would like to thank the National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands, for their contribution and ongoing support to the EPIC Study. The authors also thank all participants in the EPIC cohort for their invaluable contribution to the study and the funds supporting them. The authors would like to express our great appreciation to Tracy Lignini for the English editing service.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.numecd.2021.09.012.

References

[1] Bi X, Henry CJ. Plasma-free amino acid profiles are predictors of cancer and diabetes development. Nutr Diabetes 2017;7(3):e249.
[2] Black AE, Bingham SA, Johansson G, Coward WA. Validation of dietary intakes of protein and energy against 24 hour urinary N and DLW energy expenditure in middle-aged women, retired men and post-obese subjects: comparisons with validation against presumed energy requirements. Eur J Clin Nutr 1997;51(6):405–13.
[3] Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8(2):135–60.
[4] Chambers SA, Rowa-Dewar N, Radley A, Dobbie F. A systematic review of grandparents' influence on grandchildren's cancer risk factors. PLoS One 2017;12(11):e0185420.
[5] Ishihara J, Todoriki H, Inoue M, Tsugane S. Validity of a self-administered food-frequency questionnaire in the estimation of amino acid intake. Br J Nutr 2009;101(9):1393–9.
[6] Jennings A, MacGregor A, Pallister T, Spector T, Cassidy A. Associations between branched chain amino acid intake and biomarkers of adiposity and cardiometabolic health independent of genetic factors: a twin study. Int J Cardiol 2016;223:992–8.

[7] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74.
[8] Miyagi Y, Higashiyama M, Gochi A, Akaike M, Ishikawa T, Miura T, et al. Plasma free amino acid profiling of five types of cancer patients and its application for early detection. PLoS One 2011;6(9):e24143.
[9] Nagata C, Nakamura K, Wada K, Tsuji M, Tamai Y, Kawachi T. Branched-chain amino acid intake and the risk of diabetes in a Japanese community: the Takayama study. Am J Epidemiol 2013; 178(8):1226–32.
[10] Ratner B. The correlation coefficient: its values range between +1/−1, or do they? 2009;17(2):139–43.
[11] Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr 2002;5(6b):1113–24.
[12] Riboli E, Kaaks R. The EPIC project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol 1997;26(Suppl 1):S6–14.
[13] Schmidt JA, Rinaldi S, Scalbert A, Ferrari P, Achaintre D, Gunter MJ, et al. Plasma concentrations and intakes of amino acids in male meat-eaters, fish-eaters, vegetarians and vegans: a cross-sectional analysis in the EPIC-Oxford cohort. Eur J Clin Nutr 2016;70(3):306–12.
[14] Scrimshaw NS. INFOODS: the international network of food data systems. Am J Clin Nutr 1997;65(4 Suppl):1190s–3s.
[15] Shim JS, Oh K, Kim HC. Dietary assessment methods in epidemiologic studies. Epidemiol Health 2014;36.
[16] Slimani N, Bingham S, Runswick S, Ferrari P, Day NE, Welch AA, et al. Group level validation of protein intakes estimated by 24-hour diet recall and dietary questionnaires against 24-hour urinary nitrogen in the European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study. Cancer Epidemiol Biomarkers Prev 2003;12(8):784–95.
[17] Slimani N, Deharveng G, Unwin I, Southgate DA, Vignat J, Skeie G, et al. The EPIC nutrient database project (ENDB): a first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. Eur J Clin Nutr 2007;61(9): 1037–56.
[18] Slimani N, Ferrari P, Ocke M, Welch A, Boeing H, Liere M, et al. 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. Eur J Clin Nutr 2000;54(12):900–17.
[19] Slimani N, Kaaks R, Ferrari P, Casagrande C, Clavel-Chapelon F, Lotze G, et al. European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study: rationale, design and population characteristics. Public Health Nutr 2002;5(6B):1125–45.
[20] US Department of Agriculture, A. R. S., Nutrient Data Laboratory. USDA National Nutrient Database for Standard Reference, Release 27 (slightly revised). 2015.
[21] USDA; US Department of Agriculture, A. R. S., Nutrient Data Laboratory. USDA National Nutrient Database for Standard Reference, Release 25-27. Version Current: May 2015. Internet: http://www.ars.usda.gov/ba/bhnrc/ndl.
[22] Yamamoto J, Nishio S, Hattanda F, Nakazawa D, Kimura T, Sata M, et al. Branched-chain amino acids enhance cyst development in autosomal dominant polycystic kidney disease. Kidney Int 2017; 92(2):377–87.