

1 **Dietary amino acids and risk of stroke subtypes: a prospective analysis of**
2 **356,000 participants in seven European countries.**

3

4 Tammy Y.N. Tong¹, Robert Clarke², Julie A. Schmidt^{1,3}, Inge Huybrechts⁴, Urwah Noor¹,
5 Nita G. Forouhi⁵, Fumiaki Imamura⁵, Ruth C. Travis¹, Elisabete Weiderpass⁶, Krasimira
6 Aleksandrova^{7,8}, Christina C. Dahm⁹, Yvonne T. van der Schouw¹⁰, Kim Overvad⁹, Cecilie
7 Kyrø¹¹, Anne Tjønneland¹¹, Rudolf Kaaks¹², Verena Katzke¹², Catarina Schiborn^{13,14},
8 Matthias B. Schulze^{13,15}, Ana-Lucia Mayen-Chacon⁴, Giovanna Masala¹⁶, Sabina Sieri¹⁷,
9 Maria Santucci de Magistris¹⁸, Rosario Tumino¹⁹, Carlotta Sacerdote²⁰, Jolanda M.A. Boer²¹,
10 W.M. Monique Verschuren^{10,21}, Magritt Brustad^{22,23}, Therese Haugdahl Nøst^{22,24}, Marta
11 Crous-Bou^{25,26}, Dafina Petrova²⁷⁻²⁹, Pilar Amiano²⁹⁻³¹, José María Huerta^{29,32}, Conchi
12 Moreno-Iribas^{29,33}, Gunnar Engström³⁴, Olle Melander^{34,35}, Kristina Johansson³⁶, Kristina
13 Lindvall³⁷, Elom K. Aglago³⁸, Alicia K. Heath³⁸, Adam S. Butterworth³⁹⁻⁴², John Danesh³⁹⁻⁴³,
14 Timothy J. Key¹

15

16 **Affiliations:**

- 17 1. Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford,
18 United Kingdom
- 19 2. Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health,
20 Oxford, United Kingdom
- 21 3. Department of Clinical Epidemiology, Department of Clinical Medicine, Aarhus University Hospital and
22 Aarhus University, Aarhus, Denmark.
- 23 4. Nutrition and Metabolism Branch, International Agency for Research on Cancer (IARC), World Health
24 Organization (WHO), Lyon, France.

- 25 5. MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, United
26 Kingdom
- 27 6. International Agency for Research on Cancer (IARC), World Health Organization (WHO), Lyon, France.
- 28 7. Department Epidemiological Methods and Etiological Research, Leibniz Institute for Prevention Research
29 and Epidemiology - BIPS, Bremen, Germany.
- 30 8. Faculty of Human and Health Sciences, University of Bremen, Grazer Straße 2, 28359, Bremen, Germany.
- 31 9. Department of Public Health, Aarhus University, Aarhus, Denmark.
- 32 10. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University,
33 Utrecht, The Netherlands.
- 34 11. Danish Cancer Society Research Center, Copenhagen, Denmark.
- 35 12. Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.
- 36 13. Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke,
37 Nuthetal, Germany.
- 38 14. German Center for Diabetes Research (DZD), Neuherberg, Germany
- 39 15. Institute for Nutritional Science, University of Potsdam, Nuthetal, Germany.
- 40 16. Institute for cancer research, prevention and clinical network (ISPRO), Florence, Italy.
- 41 17. Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan,
42 Italy
- 43 18. Azienda Ospedaliera Universitaria Federico II, Napoli, Italy.
- 44 19. Hyblean Association for Epidemiological Research AIRE-ONLUS, Ragusa, Italy.
- 45 20. Unit of Cancer Epidemiology, Città della Salute e della Scienza University-Hospital, Turin, Italy.
- 46 21. National Institute for Public Health and the Environment, Bilthoven, the Netherlands.
- 47 22. Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway,
48 Tromsø, Norway.
- 49 23. The Public Dental Service Competence Centre of Northern Norway (TkNN), Tromsø, Norway.
- 50 24. K.G. Jebsen Centre for Genetic Epidemiology, Department of Public Health and Nursing, Norwegian
51 University of Science and Technology, Trondheim, Norway,
- 52 25. Unit of Nutrition and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology
53 (ICO) - Bellvitge Biomedical Research Institute (IDIBELL). L'Hospitalet de Llobregat, Barcelona, Spain.
- 54 26. Department of Epidemiology, Harvard T.H. Chan School of Public Health. Boston, MA, USA.

- 55 27. Escuela Andaluza de Salud Pública (EASP), Granada, Spain
- 56 28. Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain
- 57 29. Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Instituto de
58 Salud Carlos III, C. de Melchor Fernández Almagro, Madrid, Spain.
- 59 30. Ministry of Health of the Basque Government, Sub Directorate for Public Health and Addictions of
60 Gipuzkoa, San Sebastian, Spain.
- 61 31. Biodonostia Health Research Institute, Epidemiology of Chronic and Communicable Diseases Group, , San
62 Sebastián, Spain.
- 63 32. Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain.
- 64 33. Navarra Institute for Health Research (IdiSNA), Pamplona, Spain Instituto de Salud Pública de Navarra,
65 IdiSNA, Navarre Institute for Health Research, Pamplona, Spain
- 66 34. Department of Clinical Science in Malmö, Lund University, Clinical Research Center, Malmö, Sweden.
- 67 35. Department of Emergency and Internal Medicine, Skåne University Hospital, Malmö, Sweden.
- 68 36. Department of Public Health and Clinical medicine, Umeå University, Umeå, Sweden
- 69 37. Department of Epidemiology and Global Health, Umeå University, Umeå, Sweden
- 70 38. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London,
71 United Kingdom.
- 72 39. British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary
73 Care, University of Cambridge, Cambridge, United Kingdom.
- 74 40. National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics,
75 University of Cambridge, Cambridge, United Kingdom.
- 76 41. British Heart Foundation Centre of Research Excellence, University of Cambridge, Hills Road, Cambridge,
77 United Kingdom.
- 78 42. Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, United
79 Kingdom.
- 80 43. Department of Human Genetics, Wellcome Sanger Institute, Hinxton, United Kingdom.
- 81
- 82 Corresponding author: Dr Tammy Y.N. Tong (tammy.tong@ndph.ox.ac.uk)
83 Nuffield Department of Population Health, University of Oxford, Richard Doll Building, Old
84 Road Campus, Oxford OX3 7LF, UK

85 **Abstract**

86 *Background:* Previously reported associations of protein-rich foods with stroke subtypes have
87 prompted interest in assessment of individual amino acids. We examined the associations of
88 dietary amino acids with risks of ischaemic and haemorrhagic stroke in the EPIC study.

89 *Methods:* We analysed data on 356,142 participants from seven European countries. Dietary
90 intakes of 19 individual amino acids were assessed using validated country-specific dietary
91 questionnaires, calibrated using additional 24-hour dietary recalls. Multivariable-adjusted
92 Cox regression models were used to estimate hazard ratios (HRs) and 95% confidence
93 intervals (CIs) of ischaemic and haemorrhagic stroke in relation to intake of each amino acid.
94 The role of blood pressure as a potential mechanism was assessed in 267,642 (75%)
95 participants.

96 *Results:* After a median follow-up of 12.9 years, 4,295 participants had an ischaemic stroke
97 and 1,375 participants had a haemorrhagic stroke. After correction for multiple testing, higher
98 intake of proline (as percent of total protein) was associated with 12% lower risk of
99 ischaemic stroke (HR per 1 SD higher intake 0.88; 95% CI 0.82, 0.94). The association
100 persisted after mutual adjustment for all other amino acids, systolic and diastolic blood
101 pressure. The inverse associations of isoleucine, leucine, valine, phenylalanine, threonine,
102 tryptophan, glutamic acid, serine and tyrosine with ischaemic stroke were each attenuated
103 with adjustment for proline intake. For haemorrhagic stroke, no statistically significant
104 associations were observed in the continuous analyses after correcting for multiple testing.

105 *Conclusion:* Higher proline intake was associated with a lower risk of ischaemic stroke,
106 independent of other dietary amino acids and blood pressure.

107 **Keywords:** amino acids, dietary protein, ischaemic stroke, haemorrhagic stroke, nutritional
108 epidemiology, prospective cohort

109 **Background**

110 While stroke is the second leading cause of death worldwide, the incidence of stroke varies
111 substantially between countries reflecting differences in risk factors including variations in
112 dietary intakes [1]. Differences in dietary protein has been suggested to be important, but
113 previous studies of the associations of dietary protein intake with risk of stroke have reported
114 conflicting results, possibly reflecting heterogeneity by pathological stroke types and dietary
115 sources of protein [2, 3]. The different sources of dietary protein reflect a different profile of
116 individual dietary amino acids. For example, meat and meat products are a main source of the
117 essential amino acids and glycine, while grain products are a main source of dietary cysteine
118 [4]. A previous prospective study reported an inverse association of dietary cysteine with risk
119 of total stroke [5], and another reported an inverse association of glutamic acid but positive
120 association of glycine with stroke mortality [6], but overall there is very limited evidence on
121 the topic. Moreover, it is unclear whether the absolute quantity or the proportions of each
122 individual dietary amino acid (i.e. protein quality) that make up total dietary protein would be
123 more relevant for health. The two main stroke subtypes (ischaemic and haemorrhagic) also
124 have different dietary [7] and non-dietary (including genetic) risk factors [8, 9], which
125 highlights the importance of examining stroke subtypes separately. For example, previous
126 analyses in EPIC (the European Prospective Investigation into Cancer and Nutrition) reported
127 inverse associations of ischaemic stroke with consumption of fruit and vegetables, dietary
128 fibre and dairy products, while the risk of haemorrhagic stroke was positively associated with
129 egg consumption [7]. Therefore, any associations with dietary amino acids are likely to vary
130 by stroke type.

131 The aim of the present study was to assess the associations of both absolute and relative
132 quantities of individual dietary amino acids with risk of stroke subtypes in a prospective
133 study of 356,000 participants from seven European countries. Additionally, we aimed to

134 assess if these associations were independent of blood pressure, the major risk factor for both
135 subtypes of stroke.

136

137 **Methods**

138 *Study population*

139 The present analyses involved 356,142 participants recruited from 20 centres between 1992
140 and 2000 in seven European countries (Denmark, Germany, Italy, the Netherlands, Spain,
141 Sweden, and the UK), who participated in the EPIC study. Details of the study design have
142 been described previously [10, 11]. The inclusion and exclusion criteria for the present study
143 are shown in Supplementary figure 1. All participants provided written informed consent, and
144 the study protocol was approved by the ethical review board of IARC and the institutions
145 where the participants were recruited [10].

146

147 *Data collection*

148 At recruitment all participants completed questionnaires on medical history and socio-
149 demographic factors, as well as validated country-specific dietary questionnaires (mostly
150 food frequency questionnaires or diet histories) which asked about diet in the previous year
151 [10]. In addition, a stratified random sample of 8% of participants across all centres also
152 completed a standardised and computerised 24-hour recall, on average ~1.4 years after
153 recruitment, which was used to calibrate the dietary exposures to reduce between-centre
154 heterogeneity and to correct for measurement error [12, 13]. The calibration method and
155 rationale have been described previously [7, 12, 13].

156 Based on reported dietary intakes in both the baseline dietary questionnaires and 24-hour
157 recalls, estimates of individual dietary amino acids were derived by matching to the National
158 Nutrient Database for Standard Reference of the United States (developed at the United
159 States Department of Agriculture, or USDA) food composition tables, and appended to the
160 EPIC Nutrient Database. The matching and validation process has been reported in detail
161 [14]. USDA estimates of dietary amino acids were used because the equivalent estimates
162 were not available from all ten national Food Composition Databases that were originally
163 used in EPIC. Altogether, both calibrated and observed estimates were available for the
164 intakes of 19 individual dietary amino acids. Of the 20 standard amino acids, glutamine and
165 asparagine were not available as these two amino acids were affected by deamination
166 reactions during the acid hydrolysis process used to compile the USDA database [15, 16].
167 Cystine (two cysteine molecules joined by a disulfide bond) [17] was estimated instead of
168 cysteine, while hydroxyproline, which is not a standard amino acid, was also estimated. The
169 sums of branched-chain amino acids (isoleucine, leucine, valine), other essential amino acids
170 (histidine, lysine, methionine, phenylalanine, threonine and tryptophan) and non-essential
171 amino acids (alanine, arginine, aspartic acid, cystine, glutamic acid, glycine, hydroxyproline,
172 proline, serine and tyrosine) were also calculated.

173 The primary outcomes were ischaemic stroke (ICD 9 433-434 or ICD 10 I63) and
174 haemorrhagic stroke (ICD 9 430-431 or ICD 10 I60-I61). In addition, we also considered
175 intracerebral haemorrhage (ICD 9 431 or ICD 10 I61) and subarachnoid haemorrhage (ICD 9
176 430 or ICD 10 I60) as separate outcomes. Both non-fatal and fatal incident events were
177 considered, and details of the ascertainment process and validation methods have been
178 previously reported [7, 11]. Details on the data collection for blood pressure (available for
179 most of the cohort) and other covariates have also been reported previously [7, 11].

180

181 *Statistical analyses*

182 Cohort characteristics and dietary intakes of amino acids at baseline were summarised as
183 means (SD), median (25th, 75th percentile) or absolute numbers (%). Spearman correlation
184 coefficients were estimated between the individual dietary amino acids, and also for dietary
185 amino acids with dietary protein and food sources of protein. Intakes of dietary amino acids
186 were expressed in three different ways, primarily as individual amino acids modelled as a
187 percentage of total protein intake (% of total protein), and secondarily as individual amino
188 acids modelled as grams per day (g/day) and grams per 1000 kcal (g/1000 kcal). The three
189 approaches tested the hypotheses that the relative protein quality or protein makeup, the
190 absolute quantity of amino acids, and the relative quantity of amino acids to overall diet were
191 relevant to risk of stroke. The primary approach of expressing amino acids as a percentage of
192 total protein was selected since this approach was less susceptible to collinearity with other
193 dietary variables [18]. Because the sum of the individual amino acids did not add up to 100%
194 of total protein (mean=86.2%, partly because glutamine and asparagine were missing), we
195 included post hoc analyses expressing the amino acids as a percentage of the sum of all
196 available amino acids. The associations of total dietary protein with stroke risk were also
197 assessed, by expressing protein in g/day and g/1000 kcal.

198 Using multivariable-adjusted Cox regression, we assessed the hazard ratios (HRs) and 95%
199 confidence intervals (CIs) for sex-specific SD differences in calibrated intakes of each dietary
200 amino acid, and sex-specific SD differences and sex-specific fifths of observed intakes. For
201 each association, a test of trend was estimated by fitting the calibrated or observed intake of
202 each amino acid as a continuous variable when modelling per SD differences, or by fitting the
203 median values of each fifth as a pseudo-continuous variable when modelling differences by
204 fifths of intake. Tests of non-linearity were performed by using a likelihood-ratio test to

205 compare models with the exposure fitted as a continuous variable versus a categorical
206 variable (fifths of intake), based on the observed intake.

207 The underlying time variable for the Cox regression was age at recruitment to age at
208 diagnosis, death or administrative censoring. All analyses were stratified by sex and EPIC
209 centre, and adjusted for age at recruitment (continuous, Model 1). Model 2 additionally
210 adjusted for total energy intake (continuous, calibrated or observed intake based on USDA
211 estimates). Model 3 additionally adjusted for smoking status and intensity (never, former,
212 current<10, 10-19, 20+ cigarettes/day, unknown), current calibrated or observed alcohol
213 consumption (non-drinker (<0.1), 0.1-4.9, 5.0-14.9, 15-29.9, 30-59.9, 60+ g/day based on
214 USDA estimates), physical activity (inactive, moderately inactive, moderately active, active,
215 unknown), employment status (employed or student, neither employed nor student, unknown)
216 and level of education completed (none or primary, secondary, vocational or university,
217 unknown). Model 4 additionally adjusted for self-reported history of diabetes (yes, no,
218 unknown), prior hypertension (yes, no, unknown) and prior hyperlipidaemia (yes, no,
219 unknown). Model 5 additionally adjusted for body mass index (<22.5, 22.5-24.9, 25.0-27.4,
220 27.5-29.9, ≥ 30.0 kg/m², unknown), and was considered the main model in these analyses. As
221 total energy intake was held constant (i.e. adjusted for) in the regression model, a higher
222 intake of any individual amino acid, which contributes to total energy, would result in a
223 simultaneously lower intake of some other energy-providing foods or nutrients [19].
224 Therefore, in model 6, we removed the adjustment for total energy intake thus allowing the
225 comparison of the results with (Model 5) and without (Model 6) a theoretical substitution
226 effect of the amino acid of interest for an unspecified energy-providing source. To further
227 assess the influence of major macronutrients and to allow the interpretation of the findings
228 when holding total protein intake constant, we fitted an additional model based on Model 6
229 (without energy intake), but with the inclusion of intakes of total protein (continuous), total

230 carbohydrates (continuous), saturated fats (continuous) and unsaturated fats (continuous).
231 Likelihood-ratio χ^2 statistics were estimated at each adjustment level (for models 1-6) by
232 comparing regression models with and without inclusion of the amino acids of interest using
233 a likelihood ratio test [20]. The proportional hazards assumption was assessed using
234 Schoenfeld residuals. The assumption was not met for two covariates in the models for
235 ischaemic stroke (history of diabetes and prior hypertension) and for haemorrhagic stroke
236 (prior hypertension and BMI). However, fitting these covariates as stratifying variables
237 fulfilled the assumption and yielded close to identical estimates as adjusting, so they were
238 retained as adjustment variables.

239 To investigate whether the observed associations for the amino acids were independent from
240 the other amino acids, we included additional models mutually adjusting for each of the other
241 dietary amino acids, one at a time. We also included post-hoc analyses adjusting the amino
242 acids that remain significant across different models for its major food sources. To assess
243 whether the associations were independent of blood pressure, we estimated the adjusted mean
244 systolic and diastolic blood pressure by fifths of each dietary amino acid in the subset of
245 participants with blood pressure measurements (n=267,642), using the same covariate
246 adjustments as above. Additionally, we repeated the Cox regression analyses in this subset
247 (based on Model 5), with and without further adjustment for systolic or diastolic blood
248 pressure.

249 As secondary analyses, we repeated the main analyses for the two subtypes of haemorrhagic
250 stroke (intracerebral haemorrhage and subarachnoid haemorrhage). To assess potential
251 heterogeneity of the associations by subgroups of the major confounders and to assess
252 residual confounding, we also examined the results stratified by age at recruitment (<55, 55-
253 64, ≥ 65 years), sex, BMI (<25, 25-29.9, ≥ 30 kg/m²), smoking status (never, former, current
254 smokers), alcohol drinking (non-drinkers, moderate drinkers <15g/day, heavy drinkers

255 $\geq 15\text{g/day}$) and history of diseases (as a dichotomy of no disease history vs history of diabetes,
256 hypertension or hyperlipidaemia). We also examined the associations based on observed
257 intakes stratified by EPIC country and pooled using a meta-analysis approach. Tests of
258 heterogeneity of trend between subgroups were conducted by comparing the risk coefficients
259 in each subgroup using inverse variance weighting, testing for statistical significance with a
260 χ^2 test on $k-1$ degrees of freedom, where k is the number of subgroups. To assess whether the
261 overall results might be influenced by reverse causation, we repeated the analyses for
262 ischaemic and haemorrhagic stroke after excluding the first four years of follow-up. A
263 complete case analysis was also performed to assess the potential influence of missing
264 covariates, which were coded as a missing category in the main analyses.

265 All tests for statistical significance were two sided. Conventional p-values are reported, but
266 results were interpreted with consideration of multiple testing. To account for multiple testing
267 while allowing for correlation between the exposures, we conducted a principal component
268 analysis of the exposure variables, and determined that the first four principal components
269 explained 99% of the total variation in the exposure data [21, 22]. Consequently, the effective
270 number of independent tests was determined to be four, and the statistical significance level
271 after Bonferroni correction for multiple testing based on this number was defined as
272 $0.05/4=0.0125$. All analyses were performed using Stata version 17.0 (StataCorp, TX, USA).
273 All figures were generated using “Jasper makes plots” package [23] version 2-266 in R
274 version 4.2.1.

275

276 **Results**

277 *Baseline and dietary characteristics*

278 The baseline characteristics of the study participants are shown in Table 1 and additionally in
279 detail in Supplementary table 1. Intakes of dietary amino acids, total protein and total energy
280 are shown by sex in Supplementary table 2, and by EPIC country in Supplementary table 3.
281 The Spearman correlation coefficients between the individual dietary amino acids are shown
282 in Supplementary figure 2, and for amino acids with dietary protein and food sources of
283 protein in Supplementary figures 3 and 4 respectively.

284

285 *Associations of dietary amino acids with stroke risk*

286 After a median follow-up of 12.9 years, there were 4,295 cases of ischaemic stroke and 1,375
287 cases of haemorrhagic stroke among 356,142 participants (4,529,626 person-years in total).
288 In the fully adjusted models, associations between dietary amino acids and risks of ischaemic
289 and haemorrhagic stroke were numerically similar irrespective of whether the amino acids
290 were expressed as a percentage of total protein (Figure 1, Supplementary tables 4-5), in g/day
291 (Supplementary figure 5, Supplementary tables 4-5), g/1000 kcal (Supplementary figure 6,
292 Supplementary tables 4-5), or as a percentage of total amino acids (Supplementary figure 7,
293 Supplementary tables 4-5). Unless otherwise stated, the primary results reported on the
294 associations that remained statistically significant after correcting for multiple testing, or
295 results from categorical analyses where we observed non-linear associations.

296 When expressing the amino acids as a percentage of total protein, in the fully adjusted models
297 (Model 5) we found inverse associations of ischaemic stroke for each 1 SD higher calibrated
298 intake of total and individual branched-chain amino acids (HR [95% CI] per SD higher of
299 total: 0.93 [0.88, 0.98], isoleucine: 0.93 [0.88, 0.98], leucine: 0.93 [0.87, 0.98], valine: 0.92
300 [0.88, 0.97]), phenylalanine (0.92 [0.87, 0.98]), threonine (0.93 [0.88, 0.99]), tryptophan
301 (0.94 [0.89, 0.99]), total non-essential amino acids (0.92 [0.87, 0.98]), glutamic acid (0.90

302 [0.84, 0.96]), proline (0.88 [0.82, 0.94]), serine (0.92 [0.88, 0.97]) and tyrosine (0.92 [0.88,
303 0.98]) (Figure 1). Glycine showed a non-linear association with ischaemic stroke in the
304 categorical analysis (HR [95% CI] 1.28 [1.14, 1.43] for top versus bottom fifth, p non-
305 linearity<0.001), but only when expressing glycine as a percentage of estimated intakes of
306 total protein or total amino acids (Supplementary table 4). For haemorrhagic stroke, no
307 statistically significant associations were observed in the continuous analyses after correcting
308 for multiple testing (Figure 1). In categorical analyses, lower risks were observed from the
309 second to top fifth of proline intake compared to bottom fifth (top versus bottom fifth: 0.72
310 [0.58, 0.90], p-trend across categories=0.006, Supplementary table 5). No statistically
311 significant associations were observed for total protein with either ischaemic or haemorrhagic
312 stroke (Supplementary figures 4-5, Supplementary tables 4-5), or in any analyses examining
313 the two haemorrhagic stroke subtypes as outcomes (Supplementary table 6). Results from the
314 different covariate adjustment models are shown in Supplementary tables 7-9.

315

316 *Mutual adjustment of dietary amino acids*

317 Results with mutual adjustment for each of the other amino acids for the ten amino acids that
318 were inversely associated with ischaemic stroke risk, as reported above, are shown in Figure
319 2 (for the five amino acids that were significantly associated in both the calibrated and
320 observed analyses) and in Supplementary figure 8 (for the five amino acids that were
321 significantly associated in the calibrated analyses only). The inverse associations for proline
322 persisted after adjustment for all other amino acids, but the associations for other amino acids
323 were attenuated upon adjustment for proline. However, the higher risk of ischaemic stroke
324 comparing top to bottom fifth of glycine intake remained with adjustment for proline (1.27
325 [1.12, 1.43], results not shown in tables). In post-hoc analyses, the inverse associations for

326 proline with ischaemic stroke remained upon mutual adjustment for its different food sources
327 (e.g. 0.89; 0.83, 0.96 after mutual adjustment for cheese, with no significant association for
328 cheese in the same model, results not shown in tables). Based on these findings, the changes
329 in the HRs and likelihood ratio χ^2 statistics for proline are also depicted graphically in Figure
330 3, by each level of covariate adjustment.

331

332 *Assessment of role of blood pressure*

333 The adjusted mean levels of systolic and diastolic blood pressure across fifths of dietary
334 amino acids are shown in Supplementary tables 10-11. Across all amino acids, higher intakes
335 were associated with higher blood pressure, though the differences were very small (up to 1
336 mmHg difference in most cases). As expected from these small differences, the results for
337 both ischaemic and haemorrhagic stroke were almost unchanged after further adjustment for
338 either systolic or diastolic blood pressure (Supplementary table 12).

339

340 *Sensitivity analyses*

341 In stratified analyses, the main findings were consistent between subgroups of age, sex, BMI,
342 smoking status, alcohol drinking and disease status for both ischaemic and haemorrhagic
343 stroke (Supplementary tables 13-24). The pooled estimates by country using a meta-analysis
344 approach were similar (Supplementary tables 25-26), and results were consistent after
345 excluding the first four years of follow-up (Supplementary table 27) or based on a complete
346 case analysis (Supplementary table 28).

347

348 **Discussion**

349 In this large prospective study, we observed inverse associations of dietary intakes of proline
350 with risk of ischaemic stroke after adjustment of potential confounders, independent of the
351 other amino acids and of total protein intake. The associations were consistent regardless of
352 how the amino acids were modelled (i.e. percentage of protein, g/day and g/1000 kcal), and
353 across all subgroup analyses. Although inverse associations with ischaemic stroke were
354 initially observed for some other amino acids, these might be explained by the high
355 correlations between these amino acids and proline, since the associations were attenuated
356 after adjustment for proline. We further observed evidence of a higher risk of ischaemic
357 stroke in individuals with the highest intakes of dietary glycine, and a lower risk of
358 haemorrhagic stroke in people with higher intakes of dietary proline, though these
359 associations differed depending on the way dietary amino acids were modelled in the
360 analyses.

361 The present analyses provide important novel evidence on the associations between dietary
362 amino acids and stroke risk, as only two prospective studies have previously published on the
363 topic [5, 6]. The Swedish Mammography Cohort, with 1751 cases of total stroke (1311
364 ischaemic and 264 haemorrhagic after 10 years), reported an inverse association between
365 dietary cysteine and the risk of total stroke, but no other associations including for proline
366 were observed after mutual adjustment of the amino acids [5]. In contrast to that study, we
367 did not find an association of cystine with either stroke subtype; however, although cystine is
368 expected to be highly correlated with cysteine, their molecular structures are different, so the
369 two sets of results are not directly comparable. The Takayama Study in Japan, with 677
370 deaths from stroke (393 ischaemic and 153 intracerebral haemorrhage after 16 years), showed
371 that glutamic acid intake was inversely associated with total stroke mortality among women,
372 while glycine intake was positively associated with ischaemic stroke mortality in men with
373 no prior history of hypertension; the study did not report on proline [6]. In our study, we also

374 observed an inverse association of glutamic acid intake with ischaemic stroke, but this
375 association was attenuated when we further adjusted for dietary proline. In accordance with
376 the Takayama Study, we also found a higher risk of ischaemic stroke for people in the highest
377 fifth of glycine intake. The differences in intakes of the amino acids across populations
378 should be considered when comparing the findings. For example, the average intakes of both
379 glutamic acid and glycine when expressed as a percentage of total protein in the EPIC study
380 is closer to the lower end of the intakes reported in the Takayama study. Overall, compared to
381 the two previous studies, the current study is over three- and ten-fold larger in case numbers.
382 We also focused on examining ischaemic and haemorrhagic stroke separately, based on prior
383 evidence of heterogeneity in risk factors by stroke subtypes. Overall, the associations of
384 dietary amino acids with haemorrhagic stroke were less clear. As the number of
385 haemorrhagic stroke cases in the study was comparatively low, further investigations in other
386 populations with higher numbers of haemorrhagic stroke are warranted.

387 Previous studies have largely attributed the observed associations of dietary amino acids with
388 stroke risk to the influences of the amino acids on blood pressure. The INTERMAP study
389 reported inverse associations of glutamic acid and to a lesser degree proline [18], but positive
390 associations of glycine intake [24], with blood pressure. Our results did not suggest an
391 important role of blood pressure in explaining any of the findings, given that we did not
392 observe any meaningful differences in blood pressure by amino acid intake, nor did we
393 observe any differences in the HRs after further adjusting for blood pressure. However, blood
394 pressure and diet in EPIC were both measured at recruitment, which does not allow the
395 evaluation of a temporal association. Additionally, the use of blood pressure measurements
396 assessed at a single time point as an indicator of long-term blood pressure is also prone to
397 random measurement error, which could result in an underestimation of its role in explaining
398 any associations between amino acid intakes and stroke risk. Furthermore, blood pressure

399 measurements in EPIC were not taken using a standardised protocol across centres [25], but
400 our analyses were stratified by centre, so any influence on the results should be small.

401 The possible biological mechanisms for the association between proline and stroke risk are
402 not yet well defined. Despite being non-essential amino acids, proline, glycine and
403 hydroxyproline are the major constituents of collagen, which accounts for about one third of
404 proteins in the human body [26]. Collagen has an important role in maintaining the structure
405 and strength of connective tissues, cartilage and blood vessels [26, 27], and studies from
406 animal models have suggested that endogenous synthesis of these amino acids may not be
407 adequate for growth or collagen production [26, 28]. Because proline is the only precursor for
408 hydroxyproline, which in turn is a precursor for glycine [26, 29], dietary intakes of proline
409 may be of particular importance for collagen synthesis.

410 In contrast to other amino acids, proline contains an imino group in its structure instead of a
411 primary amino group, which results in its distinctive cyclic structure and exclusion from
412 standard amino acid metabolism [30, 31]. For this reason, the family of proline-containing
413 peptides, collectively called glycoprolines, cannot be degraded by most peptidases owing to
414 the presence of proline in their structures, and thus are more likely to enter the bloodstream
415 and exert various biological effects, including possible atheroprotective or neuroprotective
416 roles, regulation of insulin-like growth factor-I (IGF-I) homeostasis and lipid metabolism
417 including lowering of total cholesterol [32]. Glycoprolines may be derived from collagen or
418 IGF-Ia study in rats has shown that oral administration of cyclic glycine-proline, a
419 glycoproline derived from IGF-I, promoted neural plasticity and remodelling in animals with
420 focal ischaemic lesions, suggesting a neuroprotective effect of cyclic glycine-proline against
421 ischaemic stroke or its recovery [33]. In human studies, both cyclic glycine-proline
422 concentrations [34] and proline concentrations [35–37] have been proposed as prognostic
423 markers of stroke recovery or stroke events. However, the directions of the changes observed

424 have been inconsistent. Other studies have also reported low correlations between dietary
425 intake and circulating levels of proline [38]. Therefore, further research into the link between
426 proline intake and the physiological effects of circulating proline, as well as the role of
427 proline in both stroke incidence and prognosis (recovery or mortality) is needed.

428 Among dietary sources, the results of the present study showed that dairy products, especially
429 cheese, was most highly correlated with proline intake. Therefore, the findings of lower risks
430 of ischaemic stroke in people with higher proline intake in the present study are consistent
431 with previous observations of lower ischaemic stroke risk in people with higher intake of
432 cheese and yogurt in the same cohort [7]. Despite being a major source of saturated fat,
433 higher intakes of dairy products have been inversely associated with cardiovascular disease
434 outcomes in a meta-analysis of prospective studies [39], suggesting that some components in
435 cheese or dairy products might be protective. For example, studies had suggested that the
436 specific saturated fatty acid isomer composition of different foods [40], or the calcium [41] or
437 probiotics content [42] of dairy foods may account for their associations with stroke risk. As
438 the observed association for proline remained upon adjustment for cheese and other food
439 sources, the findings of the present study suggest that the amino acid profiles of dairy
440 products and their associations with CVD warrant further research. Meanwhile, dietary intake
441 of glycine was strongly correlated with red meat intake, which was previously found to be
442 modestly but positively associated with ischaemic stroke risk in EPIC, though this association
443 was attenuated after adjustment for fibre intake [7]. Further population-based and mechanistic
444 studies are needed to replicate the observed associations between dietary amino acids and
445 risks of stroke subtypes, prior to investigating any implications for potential dietary
446 interventions.

447 The chief strengths of this study include the large sample size with participants recruited from
448 seven European countries, the prospective design and prolonged duration of follow-up. EPIC

449 is also a well-characterised cohort with extensive dietary and lifestyle/health behavioural
450 data, allowing us to examine estimated intakes of 19 dietary amino acids with stroke risk,
451 while adjusting for multiple confounders. The ascertainment of stroke cases in this cohort
452 was either completely or partially validated depending on EPIC country, and previous dietary
453 analyses have found no heterogeneity in the associations by the extent of stroke validation
454 [7]. To adjust for potential bias due to differences between the country-specific dietary
455 questionnaires, we also applied calibration to the dietary data using additional standardised
456 24-hour recalls to reduce between-centre heterogeneity and to reduce measurement error. The
457 study also had several limitations. Diet was ascertained using a single dietary questionnaire
458 collected at baseline, which meant that we were not able to account for dietary changes
459 during follow-up. We were unable to assess the associations of dietary glutamine, asparagine
460 or cysteine with stroke risk, as estimates of these amino acids were not available from the
461 USDA database. The matching of amino acids data based on a single US database in different
462 countries across Europe may also be a limitation, and the absolute validity of the amino acid
463 estimates assessed using FFQs and 24-hour recalls is not known, although good agreement
464 and thus relative validity has been shown for estimates of total energy and protein intakes
465 between the USDA and country-specific databases in the previous validation study [14].
466 Because of the large number of tests conducted, false positive findings due to chance are
467 possible, but we focused on the findings that survived correction for multiple testing. We also
468 conducted analyses further adjusting for each one of the other amino acids, which given their
469 high correlations might constitute over-adjustment, but this approach also helped highlight
470 the clear associations with proline. As with any observational study, the present study could
471 not establish causality, and cannot fully exclude residual confounding or reverse causation.
472 However, we assessed the former by comparing the changes in χ^2 statistics across models,
473 and limited the latter possibility by our sensitivity analysis excluding the first four years of

474 follow-up. Information on other potential mediators such as cholesterol concentrations was
475 only available in a small subset of participants at study baseline, so we were unable to access
476 their potential effects. Lastly, the present study was largely restricted to white European
477 individuals, therefore the findings may not be generalisable to other populations which have
478 different sources of protein intake and dietary amino acid profiles, as well as genetic
479 differences.

480

481 **Conclusions**

482 This large study demonstrated an inverse association of both absolute intake and relative
483 proportions of dietary proline with risk of ischaemic stroke, independent of the other amino
484 acids. The suggestive associations of lower risk of haemorrhagic stroke in people with higher
485 proline intake, and higher risk of ischaemic stroke in people with higher proportions of
486 glycine intake, also warrant further investigation. These observations were largely
487 unexplained by differences in blood pressure. Further research is needed from both large-
488 scale population-based studies and mechanistic studies to replicate these findings, in addition
489 to genetic studies and intervention trials to assess the causal relevance of these associations
490 and possible implications for prevention.

491 *Acknowledgments:* The authors thank all participants in the EPIC cohort for their invaluable
492 contribution to the study.

493 *Authors' contributions:* The study was conceived and designed by TYNT and TJK with input
494 from all other authors. TYNT analysed the data and wrote the first draft of the manuscript,
495 with input from TJK, RC, JAS, IH, UN, NGF, FI, RCT, EW, KA, CCD and YTVS. All other
496 authors provided the data and revised the manuscript critically for important intellectual
497 content. TYNT had primary responsibility for final content. All authors read and approved the
498 final manuscript.

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

503 *Conflicts of interest:* The authors have no conflicts of interest to declare.

504 *Funding:* This work was supported by a Nuffield Department of Population Health
505 Fellowship (TYNT), the UK Medical Research Council (MR/M012190/1), Cancer Research
506 UK (C8221/A19170 and 570/A16491), and the Wellcome Trust (Our Planet Our Health,
507 Livestock Environment and People 205212/Z/16/Z). EPIC-CVD has been supported by the
508 European Union Framework 7 (HEALTH-F2-2012-279233), the European Research Council
509 (268834), the UK Medical Research Council (G0800270 and MR/L003120/1), the British
510 Heart Foundation (SP/09/002 and RG/08/014 and RG13/13/30194), and the UK National
511 Institute of Health Research. The establishment of the study subcohort was supported by the
512 EU Sixth Framework Programme (FP6) (grant LSHM_CT_2006_037197 to the InterAct
513 project) and the Medical Research Council Epidemiology Unit (grants MC_UU_12015/1 and
514 MC_UU_12015/5). The coordination of the European Prospective Investigation into Cancer

515 and Nutrition (EPIC) is financially supported by International Agency for Research on
516 Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of
517 Public Health, Imperial College London which has additional infrastructure support provided
518 by the NIHR Imperial Biomedical Research Centre (BRC). The national cohorts are
519 supported by: Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave
520 Roussy, Mutuelle Générale de l'Éducation Nationale, Institut National de la Santé et de la
521 Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research
522 Center (DKFZ), German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), Federal
523 Ministry of Education and Research (BMBF) (Germany); Associazione Italiana per la
524 Ricerca sul Cancro-AIRC-Italy, Compagnia di SanPaolo and National Research Council
525 (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), LK Research Funds,
526 Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research
527 Fund (WCRF) (The Netherlands), and Statistics Netherlands is acknowledged for providing
528 the causes of death; Health Research Fund (FIS) - Instituto de Salud Carlos III (ISCIII),
529 Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, and the
530 Catalan Institute of Oncology - ICO (Spain); Swedish Cancer Society, Swedish Research
531 Council and County Councils of Skåne and Västerbotten and the funds supporting the
532 Northern Sweden Diet Database (Sweden); Cancer Research UK (14136 to EPIC-Norfolk;
533 C8221/A29017 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk;
534 MR/M012190/1 to EPIC-Oxford) (United Kingdom). The funders had no role in the design
535 of the study; in the collection, analyses, or interpretation of data; in the writing of the
536 manuscript; or in the decision to publish the results.

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

539 **References**

- 540 1. Feigin VL, Stark BA, Johnson CO, et al (2021) Global, regional, and national burden
541 of stroke and its risk factors, 1990-2019: A systematic analysis for the Global Burden
542 of Disease Study 2019. *Lancet Neurol* 20:1–26. [https://doi.org/10.1016/S1474-4422\(21\)00252-0](https://doi.org/10.1016/S1474-4422(21)00252-0)
543
- 544 2. Zhang Z, Xu G, Yang F, et al (2014) Quantitative analysis of dietary protein intake and
545 stroke risk. *Neurology* 83:19–25. <https://doi.org/10.1212/WNL.0000000000000551>
- 546 3. Zhang XW, Yang Z, Li M, et al (2016) Association between dietary protein intake and
547 risk of stroke: A meta-analysis of prospective studies. *Int J Cardiol* 223:548–551.
548 <https://doi.org/10.1016/j.ijcard.2016.08.106>
- 549 4. Górska-Warsewicz H, Laskowski W, Kulykovets O, et al (2018) Food Products as
550 Sources of Protein and Amino Acids-The Case of Poland. *Nutrients* 10:.
551 <https://doi.org/10.3390/nu10121977>
- 552 5. Larsson SC, Håkansson N, Wolk A (2015) Dietary Cysteine and Other Amino Acids
553 and Stroke Incidence in Women. *Stroke* 46:922–6.
554 <https://doi.org/10.1161/STROKEAHA.114.008022>
- 555 6. Nagata C, Wada K, Tamura T, et al (2015) Dietary intakes of glutamic acid and
556 glycine are associated with stroke mortality in Japanese adults. *J Nutr* 145:720–728.
557 <https://doi.org/10.3945/jn.114.201293>
- 558 7. Tong TYN, Appleby PN, Key TJ, et al (2020) The associations of major foods and
559 fibre with risks of ischaemic and haemorrhagic stroke: a prospective study of 418 329
560 participants in the EPIC cohort across nine European countries. *Eur Heart J* 44:1–11.
561 <https://doi.org/10.1093/eurheartj/ehaa007>
- 562 8. Sun L, Clarke R, Bennett D, et al (2019) Causal associations of blood lipids with risk
563 of ischemic stroke and intracerebral hemorrhage in Chinese adults. *Nat Med* 25:569–
564 574. <https://doi.org/10.1038/s41591-019-0366-x>
- 565 9. Price AJ, Wright FL, Green J, et al (2018) Differences in risk factors for 3 types of
566 stroke: UK prospective study and meta-analyses. *Neurology* 90:e298–e306.
567 <https://doi.org/10.1212/WNL.0000000000004856>
- 568 10. Riboli E, Hunt K, Slimani N, et al (2002) European Prospective Investigation into
569 Cancer and Nutrition (EPIC): study populations and data collection. *Public Health*
570 *Nutr* 5:1113. <https://doi.org/10.1079/PHN2002394>
- 571 11. Danesh J, Saracci R, Berglund G, et al (2007) EPIC-Heart: The cardiovascular
572 component of a prospective study of nutritional, lifestyle and biological factors in
573 520,000 middle-aged participants from 10 European countries. *Eur J Epidemiol*
574 22:129–141. <https://doi.org/10.1007/s10654-006-9096-8>
- 575 12. Slimani N, Ferrari P, Ocké M, et al (2000) Standardization of the 24-hour diet recall
576 calibration method used in the European Prospective Investigation into Cancer and
577 Nutrition (EPIC): general concepts and preliminary results. *Eur J Clin Nutr* 54:900–17
- 578 13. Ferrari P, Day NE, Boshuizen HC, et al (2008) The evaluation of the diet/disease
579 relation in the EPIC study: considerations for the calibration and the disease models.
580 *Int J Epidemiol* 37:368–378. <https://doi.org/10.1093/ije/dym242>

- 581 14. Iguacel I, Perez-Cornago A, Schmidt JA, et al (2022) Evaluation of protein and amino
582 acid intake estimates from the EPIC dietary questionnaires and 24-h dietary recalls
583 using different food composition databases. *Nutr Metab Cardiovasc Dis* 32:80–89.
584 <https://doi.org/10.1016/j.numecd.2021.09.012>
- 585 15. Lenders CM, Liu S, Wilmore DW, et al (2009) Evaluation of a novel food composition
586 database that includes glutamine and other amino acids derived from gene sequencing
587 data. *Eur J Clin Nutr* 63:1433–1439. <https://doi.org/10.1038/ejcn.2009.110>
- 588 16. Mustănea G, Ungureanu EL, Iorga E (2019) Protein acidic hydrolysis for amino acids
589 analysis in food - progress over time: A short review. *J Hyg Eng Des* 26:81–87
- 590 17. National Center for Biotechnology Information (2022) PubChem Compound Summary
591 for CID 595, Cystine. <https://pubchem.ncbi.nlm.nih.gov/compound/Cystine>. Accessed
592 11 Feb 2022
- 593 18. Stamler J, Brown IJ, Daviglus ML, et al (2009) Glutamic acid, the main dietary amino
594 acid, and blood pressure: The intermap study (international collaborative study of
595 macronutrients, micronutrients and blood pressure). *Circulation* 120:221–228.
596 <https://doi.org/10.1161/CIRCULATIONAHA.108.839241>
- 597 19. Ibsen DB, Laursen ASD, Würtz AML, et al (2021) Food substitution models for
598 nutritional epidemiology. *Am J Clin Nutr* 113:294–303.
599 <https://doi.org/10.1093/ajcn/nqaa315>
- 600 20. Floud S, Balkwill A, Moser K, et al (2016) The role of health-related behavioural
601 factors in accounting for inequalities in coronary heart disease risk by education and
602 area deprivation: prospective study of 1.2 million UK women. *BMC Med* 14:145.
603 <https://doi.org/10.1186/s12916-016-0687-2>
- 604 21. Gao X, Starmer J, Martin ER (2008) A multiple testing correction method for genetic
605 association studies using correlated single nucleotide polymorphisms. *Genet
606 Epidemiol* 32:361–369. <https://doi.org/10.1002/gepi.20310>
- 607 22. Adams CD, Richmond R, Santos Ferreira DL, et al (2019) Circulating metabolic
608 biomarkers of screen-detected prostate cancer in the ProtecT study. *Cancer Epidemiol
609 Biomarkers Prev* 28:208–216. <https://doi.org/10.1158/1055-9965.EPI-18-0079>
- 610 23. Arnold M (2020) Jasper: Jasper makes plots. 2020.
611 <https://github.com/arnhew99/Jasper>
- 612 24. Stamler J, Brown IJ, Daviglus ML, et al (2013) Dietary glycine and blood pressure:
613 The international study on macro/micronutrients and blood pressure. *Am J Clin Nutr*
614 98:136–145. <https://doi.org/10.3945/ajcn.112.043000>
- 615 25. Schulze MB, Kroke A, Saracci R, Boeing H (2002) The effect of differences in
616 measurement procedure on the comparability of blood pressure estimates in multi-
617 centre studies. *Blood Press Monit* 7:95–104. [https://doi.org/10.1097/00126097-
618 200204000-00002](https://doi.org/10.1097/00126097-200204000-00002)
- 619 26. Li P, Wu G (2018) Roles of dietary glycine, proline, and hydroxyproline in collagen
620 synthesis and animal growth. *Amino Acids* 50:29–38. [https://doi.org/10.1007/s00726-
621 017-2490-6](https://doi.org/10.1007/s00726-017-2490-6)
- 622 27. Xu J, Shi G-P (2014) Vascular wall extracellular matrix proteins and vascular diseases.
623 *Biochim Biophys Acta* 1842:2106–2119. <https://doi.org/10.1016/j.bbadis.2014.07.008>

- 624 28. Hou Y, Yao K, Yin Y, Wu G (2016) Endogenous synthesis of amino acids limits
625 growth, lactation, and reproduction in animals. *Adv Nutr* 7:331–342.
626 <https://doi.org/10.3945/an.115.010850>
- 627 29. Wu G, Bazer FW, Burghardt RC, et al (2011) Proline and hydroxyproline metabolism:
628 Implications for animal and human nutrition. *Amino Acids* 40:1053–1063.
629 <https://doi.org/10.1007/s00726-010-0715-z>
- 630 30. Phang JM, Donald SP, Pandhare J, Liu Y (2008) The metabolism of proline, a stress
631 substrate, modulates carcinogenic pathways. *Amino Acids* 35:681–690.
632 <https://doi.org/10.1007/s00726-008-0063-4>
- 633 31. Wyse ATS, Netto CA (2011) Behavioral and neurochemical effects of proline. *Metab*
634 *Brain Dis* 26:159–172. <https://doi.org/10.1007/s11011-011-9246-x>
- 635 32. Misiura M, Milyk W (2019) Proline-containing peptides-New insight and
636 implications: A Review. *Biofactors* 45:857–866. <https://doi.org/10.1002/biof.1554>
- 637 33. Kaneko H, Namihira M, Yamamoto S, et al (2022) Oral administration of cyclic
638 glycyl-proline facilitates task learning in a rat stroke model. *Behav Brain Res*
639 417:113561. <https://doi.org/10.1016/j.bbr.2021.113561>
- 640 34. Fan D, Krishnamurthi R, Harris P, et al (2019) Plasma cyclic glycine proline/IGF-1
641 ratio predicts clinical outcome and recovery in stroke patients. *Ann Clin Transl Neurol*
642 6:669–677. <https://doi.org/10.1002/acn3.743>
- 643 35. Goulart VAM, Sena MM, Mendes TO, et al (2019) Amino acid biosignature in plasma
644 among ischemic stroke subtypes. *Biomed Res Int* 2019:.
645 <https://doi.org/10.1155/2019/8480468>
- 646 36. Szpetnar M, Hordyjewska A, Malinowska I, et al (2016) The fluctuation of free amino
647 acids in serum during acute ischemic stroke. *Curr Issues Pharm Med Sci* 29:151–154.
648 <https://doi.org/10.1515/cipms-2016-0031>
- 649 37. Wang X, Zhang L, Sun W, et al (2021) Changes of Metabolites in Acute Ischemic
650 Stroke and Its Subtypes. *Front Neurosci* 14:1–8.
651 <https://doi.org/10.3389/fnins.2020.580929>
- 652 38. Iguacel I, Schmidt JA, Perez-Cornago A, et al (2021) Associations between dietary
653 amino acid intakes and blood concentration levels. *Clin Nutr* 40:3772–3779.
654 <https://doi.org/10.1016/j.clnu.2021.04.036>
- 655 39. Alexander DD, Bylsma LC, Vargas AJ, et al (2016) Dairy consumption and CVD: a
656 systematic review and meta-analysis. *Br J Nutr* 115:737–750.
657 <https://doi.org/10.1017/S0007114515005000>
- 658 40. Steur M, Johnson L, Sharp SJ, et al (2021) Dietary fatty acids, macronutrient
659 substitutions, food sources and incidence of coronary heart disease: Findings from the
660 EPIC-CVD case-cohort study across nine european countries. *J Am Heart Assoc* 10:.
661 <https://doi.org/10.1161/JAHA.120.019814>
- 662 41. Lorenzen JK, Astrup A (2011) Dairy calcium intake modifies responsiveness of fat
663 metabolism and blood lipids to a high-fat diet. *Br J Nutr* 105:1823–1831.
664 <https://doi.org/10.1017/S0007114510005581>
- 665 42. Mozaffarian D, Wu JHY (2018) Flavonoids, dairy foods, and cardiovascular and

666 metabolic health: A review of emerging biologic pathways. *Circ Res* 122:369–384.
667 <https://doi.org/10.1161/CIRCRESAHA.117.309008>
668

Table 1: Participant characteristics at recruitment by sex in the EPIC study

Participant characteristics	Men	Women
Number of participants	129,005	227,137
Age, years, mean (SD)	52.1 (9.9)	50.6 (10.8)
Current smokers, n (%)	36,561 (28.3%)	48,287 (21.3%)
Moderately active or active, n (%)	63,284 (49.1%)	91,104 (40.1%)
Employed or student, n (%)	85,841 (66.5%)	124,753 (54.9%)
Vocational or university degree, n (%)	67,344 (52.2%)	109,160 (48.1%)
Alcohol consumption ≥ 15.0 g/day, n (%)	60,989 (47.2%)	42,452 (18.7%)
Alcohol consumption, g/day, median (25 th , 75 th percentile)	13.8 (4.5, 31.4)	4.0 (0.6, 12.4)
Overweight or obese (BMI ≥ 25.0 kg/m ²), n (%)	81,921 (63.5%)	107,011 (47.1%)
BMI, kg/m ² , mean (SD)	26.5 (3.6)	25.5 (4.5)
History of diabetes, n (%)	3,981 (3.1%)	4,668 (2.1%)
History of hypertension, n (%)	24,828 (19.2%)	41,442 (18.2%)
History of hyperlipidaemia, n (%)	21,093 (16.4%)	24,149 (10.6%)
Total protein intake, g/day, mean (SD)	90.0 (26.6)	74.9 (21.4)
Total energy intake, kcal/day, mean (SD)	2444.5 (634.8)	1965.9 (517.4)

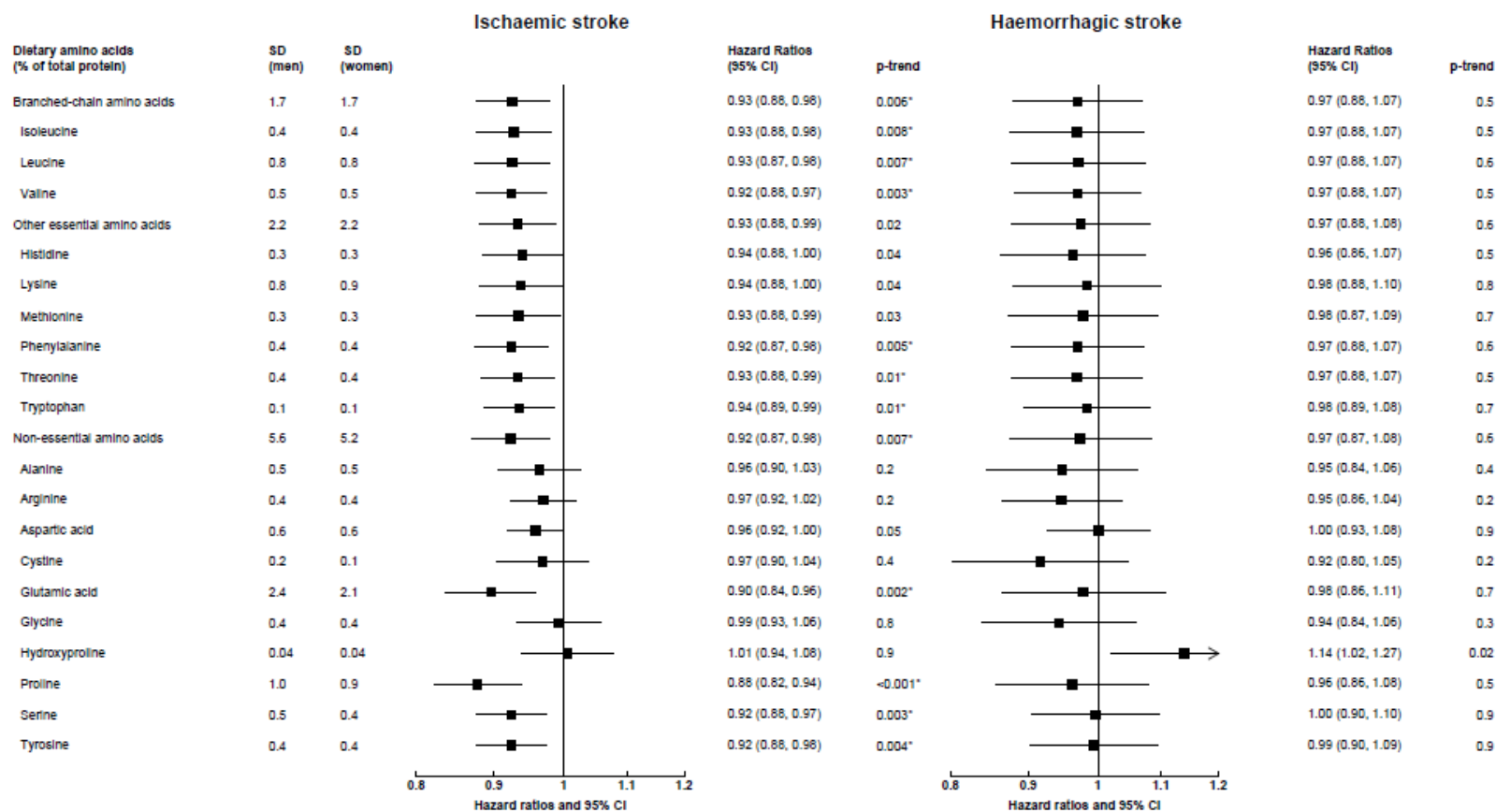


Figure 1 Hazard ratios (95% confidence intervals) for ischaemic (4295 cases) and haemorrhagic stroke (1375 cases) by increments of calibrated intakes of dietary amino acids (as percent of total protein).

Hazard ratios modelled per 1 sex-specific SD increment in dietary amino acids, expressed as percent of total protein. The model was stratified by sex and centre, and adjusted for age, calibrated energy intake, smoking, calibrated alcohol consumption, physical activity, employment status, highest level of education completed, history of diabetes, prior hypertension, prior hyperlipidaemia, body mass index. Estimates with an asterisk denote ones which were statistically significant after correcting for multiple testing.

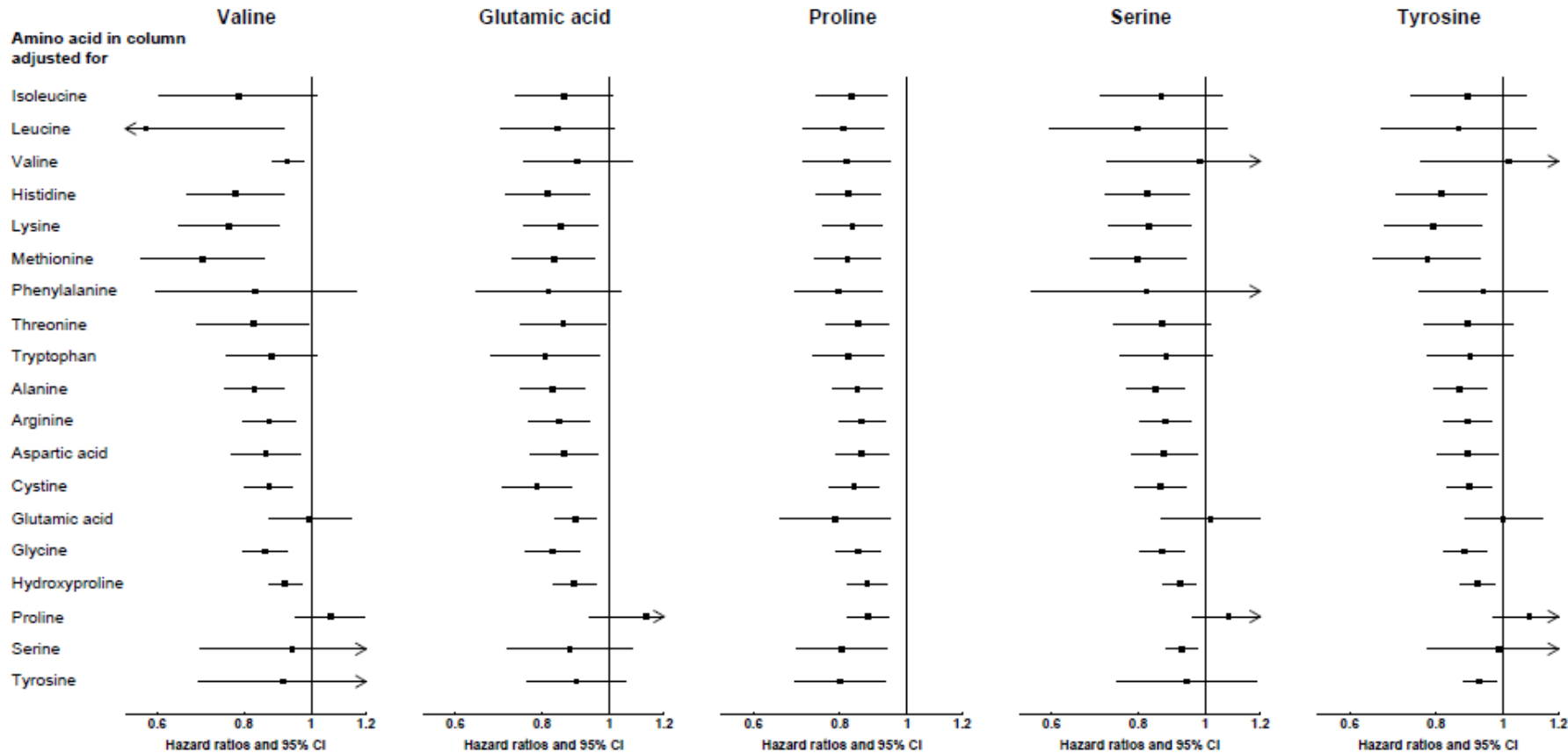


Figure 2 Hazard ratios (95% confidence intervals) for the association between selected dietary amino acids and ischaemic stroke (4295 cases), with mutual adjustment for each of the other amino acids.

Results are shown for amino acids which were statistically significantly associated with ischemic stroke in both calibrated and uncalibrated analyses. Hazard ratios modelled per 1 sex-spe SD (as shown in Figure 1) increment in dietary amino acids in column heading, expressed as percent of total protein, with multivariable adjustment plus amino acid on the left panel. If the amino acid in column and in left panel are the same, the result is interpretable as the hazard ratio for the amino acid based on the multivariable adjusted model.

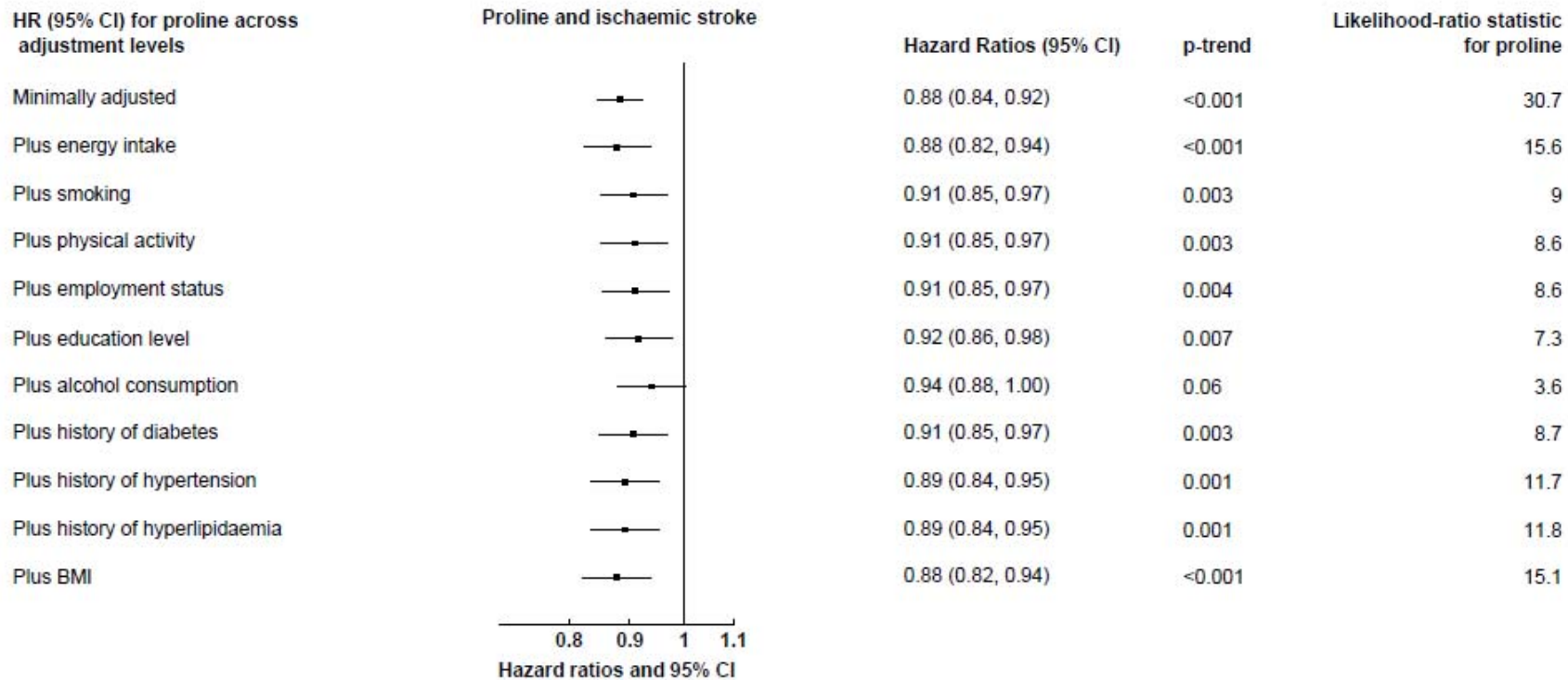


Figure 3 Hazard ratios (95% confidence intervals) for the association between dietary proline and ischaemic stroke (4295 cases), with additional adjustment each covariate.

Hazard ratios modelled per 1 SD increment (1 SD=1.0 in men, 0.9 in women) in calibrated proline intake, expressed as percent of total protein. The minimally adjusted model was stratified by sex and centre, and adjusted for age (continuous). The likelihood-ratio test statistic for proline is the χ^2 statistic based on a likelihood ratio test comparing models with and without proline at the varying levels of covariate adjustment. The changes in the χ^2 statistic across models could be interpreted as a measure of the extent to which the covariates could account for any associations between proline and ischaemic stroke risk.