1 **Plasma concentrations of advanced glycation end-products and colorectal cancer risk in the**

2 **EPIC study**

3 **Short title:** AGEs and colorectal cancer

4

5 Elom K. Aglago[1], Casper G. Schalkwijk[2], Heinz Freisling[1], Veronika Fedirko[3], David J. Hughes[4], Li

6 Jiao[5], Christina C Dahm[6], Anja Olsen[6,7], Anne Tjønneland[7,8], Verena Katzke[9], Theron Johnson[9],

7 Matthias B. Schulze[10,11], Krasimira Aleksandrova[11,12], Giovanna Masala[13], Sabina Sieri[14], Vittorio

8 Simeon[15], Rosario Tumino[16], Alessandra Macciotta[17], Bas Bueno-de-Mesquita[18], Guri Skeie[19], Inger

9 Torhild Gram[19], Torkjel Sandanger[19], Paula Jakszyn[20,21], Maria-Jose Sánchez[22, 23, 24, 25], Pilar

10 Amiano[25,26], Sandra M. Colorado-Yohar[25,27,28], Aurelio Barricarte Gurrea[25, 29, 30], Aurora Perez-

11 Cornago[31], Ana-Lucia Mayén[1], Elisabete Weiderpass[32], Marc J. Gunter[1], Alicia K. Heath[33], Mazda

12 Jenab[1]

13

14 [1]Nutrition and Metabolism Section, International Agency for Research on Cancer, Lyon, France

15 [2]Department of Internal Medicine, CARIM School for Cardiovascular Diseases, Maastricht

16 University Medical Center, The Netherlands

17 [3]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA

18 [4]Cancer Biology and Therapeutics Group (CBT), Conway Institute, School of Biomolecular and

19 Biomedical Science (SBBS), University College Dublin, Ireland

20 [5]Department of Medicine, Baylor College of Medicine, Houston, TX, USA

21 [6]Department of Public Health, Aarhus University, Denmark

22 [7]Danish Cancer Society Research Center, Denmark

23 [8]Department of Public Health, University of Copenhagen, Denmark

24 [9]Division of Cancer Epidemiology, German Cancer research Center (DKFZ), Heidelberg, Germany

25 [10]Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-

26 Rehbruecke, Nuthetal, Germany

27 [11]Institute of Nutrition Science, University of Potsdam, Nuthetal, Germany

28 [12]Nutrition, Immunity and Metabolism Department of Nutrition and Gerontology, German Institute of

29 Human Nutrition Potsdam-Rehbruecke (DIfE), Nuthetal, Germany

30 [13]Cancer Risk Factors and Life-Style Epidemiology Unit, Institute for Cancer Research, Prevention

31 and Clinical Network -ISPRO, Florence, Italy

32 [14]Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano,

33 Milano, Italy

34 [15]Dipartimento di Salute Mentale e Fisica e Medicina Preventiva, University 'Luigi Vanvitelli',

35 Napoli, Italy

36 [16]Cancer Registry and Histopathology Department, Provincial Health Authority (ASP 7), Ragusa,

37 Italy

38   [17]Department of Clinical and Biological Sciences, University of Turin, Italy

39   [18]Former senior scientist, Dept. for Determinants of Chronic Diseases (DCD), National Institute for

40   Public Health and the Environment (RIVM), PO Box 1, 3720 BA Bilthoven, The Netherlands

41   [19]Faculty of Health Sciences, Department of Community Medicine, University of Tromsø, The Arctic

42   University of Norway

43   [20]Unit of Nutrition and Cancer, Cancer Epidemiology Research Programme, Catalan Institute of

44   Oncology, Barcelona, Spain

45   [21]Blanquerna School of Health Sciences, Ramon Llull University, Barcelona, Spain

46   [22] Escuela Andaluza de Salud Pública (EASP), Granada, Spain

47   [23]Instituto de Investigación Biosanitaria ibs.Granada, Granada, Spain

48   [24]Department of Preventive Medicine and Public Health, University of Granada, Granada, Spain.

49   [25]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid,

50   Spain

51   [26]Public Health Division of Gipuzkoa, BioDonostia Research Institute, Donostia-San Sebastian, Spain

52   [27]Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain.

53   [28]Research Group on Demography and Health, National Faculty of Public Health, University of

54   Antioquia, Medellín, Colombia

55   [29]Navarra Public Health Institute, Pamplona, Spain.

56   [30]Navarra Institute for Health Research (IdiSNA) Pamplona, Spain.

57   [31]Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, UK

58   [32]Office of the Director, International Agency for Research on Cancer (IARC), Lyon, France

59   [33]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London,

60   UK

61

62   *** To whom correspondence should be addressed.:** Mazda Jenab, Address: 150 Cours Albert

63   Thomas, 69372 Lyon Cedex 08, Email: jenabm@iarc.fr, Tel: +33 472 73 80 82

64

72   **Abbreviations used:** AGE, Advanced glycation end-product; BMI, body mass index; CEL, N^ε-

73   (carboxyethyl)lysine; CI, confidence interval; CML, N^ε-(carboxymethyl)lysine; CRC, colorectal

74   cancer; EPIC, European Prospective Investigation into Cancer and Nutrition; IARC, International

75   Agency for Research on Cancer; ICD, international classification of diseases; MG-H1, Nδ-(5-hydro-

76   5-methyl-4-imidazolon-2-yl)-ornithine; RAGE, receptor for AGE; SD, standard deviation; UPLC-

77   MS/MS, ultra-performance liquid chromatography tandem mass-spectrometry

78

Abstract

Advanced glycation end-products (AGEs) are a heterogeneous group of compounds formed by the non-enzymatic reaction between amino-acids and reducing sugars, or dicarbonyls as intermediate compounds. Experimental studies suggest that AGEs may promote colorectal cancer, but prospective epidemiologic studies are inconclusive. We conducted a case-control study nested within a large European cohort. Plasma concentrations of three protein-bound AGEs: N^ε-(carboxy-methyl)lysine (CML), N^ε-(carboxy-ethyl)lysine (CEL) and N^δ-(5-hydro-5-methyl-4-imidazolon-2-yl)-ornithine (MG-H1) were measured by ultra-performance liquid chromatography tandem mass-spectrometry in baseline samples collected from 1,378 incident primary colorectal cancer cases and 1,378 matched controls. Multivariable-adjusted odds ratios (ORs) and 95% confidence intervals (CIs) were computed using conditional logistic regression for colorectal cancer risk associated with CML, CEL, MG-H1, total AGEs, and [CEL+MG-H1: CML] and [CEL:MG-H1] ratios. Inverse colorectal cancer risk associations were observed for CML (OR comparing highest to lowest quintile, $OR_{Q5vs.Q1}$=0.40, 95%CI:0.27-0.59), MG-H1 ($OR_{Q5vs.Q1}$=0.73, 95%CI:0.53 - 1.00) and total AGEs (OR $_{Q5vs.Q1}$=0.52, 95%CI:0.37 - 0.73) whereas no association was observed for CEL. A higher [CEL+MG-H1: CML] ratio was associated with colorectal cancer risk ($OR_{Q5vs.Q1}$=1.91, 95%CI:1.31-2.79). The associations observed did not differ by sex, or by tumour anatomical subsite. Although individual AGEs concentrations appear to be inversely associated with colorectal cancer risk, a higher ratio of methylglyoxal-derived AGEs versus those derived from glyoxal (calculated by [CEL+MG-H1: CML] ratio) showed a strong positive risk association. Further insight on the metabolism of AGEs and their dicarbonyls precursors, and their roles in colorectal cancer development is needed.

Summary: We evaluated colorectal cancer risk associated with blood levels of major advanced glycation end-products (AGEs). The AGEs examined were mostly associated with lower colorectal cancer risk. The ratio of methylglyoxal-derived versus glyoxal-derived AGEs was positively associated with colorectal cancer.

Introduction

Colorectal cancer (CRC) is the third most common incident cancer and the second leading cause of cancer death globally (1). A substantial body of epidemiologic evidence, particularly from large-scale prospective cohort studies, apportions a considerable contribution of modifiable dietary and lifestyle risk factors to CRC development (2-4). Western-type diets tend to promote the formation of Advanced Glycation End-products (AGEs), a heterogeneous class of pro-inflammatory and pro-oxidative compounds formed irreversibly by the non-enzymatic combination of amino acids and reducing sugars (5-8). AGEs can also be formed when proteins are glycated by highly reactive dicarbonyls such as glyoxal (GO) and methylglyoxal (MGO) absorbed from the diet, and/or smoking or produced as sugar and lipid metabolism by-products (8,9). GO and MGO have been reported to be over 20,000 times more potent in glycating amino acids, compared to sugars (10,11). As a consequence, most abundant AGEs in the body are derived from GO (N^ε-(carboxymethyl)lysine; CML) or MGO (N^ε-(carboxyethyllysine), CEL; and N^δ-(5-hydro-5-methyl-4-imidazolon-2-yl)-ornithine, MG-H1) (12-14).

AGEs are thought to affect CRC development by promoting a pro-inflammatory and oxidative environment, primarily via binding to the receptor for AGEs (RAGE), a transmembrane protein that belongs to the immunoglobulin superfamily (15). Immunohistochemical expression of AGEs is higher in colon cancer tumours compared to adjacent normal tissues and AGEs have been shown to enhance and promote colon cancer growth in *in vitro* models (16-18). Animal studies show that AGEs can induce sustained inflammation in the colon and promote colon cancer development (19,20). However, two case-control studies nested within prospective studies have reported inconclusive findings for the association between circulating AGEs levels and CRC. In the Women's Health Initiative (WHI) study, Chen et al. (21) found an inverse association between serum CML and CRC in women (Odds ratio (OR)= 0.85, 95%CI:0.49–1.47) while Jiao et al. (22) reported a positive association between circulating CML and CRC risk in male smokers (OR=1.20, 95% confidence interval CI=0.64-2.26) in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) study. These previous investigations focused on CML only, did not include other major AGEs or assess

5

134 possible differences by sex or tumour anatomical subsite. Furthermore, they detected AGEs by ELISA

135 kits, which have low specificity and reproducibility (23).

136   The aim of the present study was to examine the associations between pre-diagnostic protein-

137 bound circulating levels of CML, CEL, and MG-H1 measured using ultra-performance liquid

138 chromatography tandem mass spectrometry (UPLC-MS/MS) and CRC risk in the European

139 Prospective Investigation into Cancer and Nutrition (EPIC) cohort. The rationale for the selection of

140 these three AGEs is threefold: they are considered as the most abundant in the body, they have very

141 well characterized chemical structures, and they are derived from specific pathways of formation with

142 CML being mainly derived from GO whereas CEL and MG-H1 are mainly derived from MGO. For

143 CML and CEL, lysine is the amino-acid of the glycation site, whereas it is arginine for MG-H1. At the

144 cellular level, the MGO-lysine adduct CEL is predominantly formed in the cytosol through the

145 glycation of cytosol proteins whereas the MGO-arginine adduct MG-H1 is equally found in cytosol,

146 histone, and mitochondria proteins (24). We hypothesised that protein-bound concentrations of these

147 AGEs would be associated with a higher CRC risk. We also examined CRC risk associated with the

148 ratios of AGEs from specific dicarbonyls similar to the [CEL:MG-H1] ratio assessed in previous

149 studies (25,26), as a potential index of the chemical origin of the AGEs. Although MGO glycates

150 amino acids to CEL and MG-H1, both these MGO-derived AGEs have different promoting factors as

151 they are produced from lysine and arginine, respectively. Thus, we applied the ratio of [CEL:MG-H1]

152 as a proxy of the potential differential glycating activities of MGO in the body.

153

154 **Materials and methods**

155 Study population and data collection

156 We conducted a nested case-control study within the European Prospective Investigation into Cancer

157 and Nutrition (EPIC) cohort, an ongoing multicentre prospective study with participants recruited

158 from 23 centres constelled in 10 European countries (Denmark, France, Germany, Greece, Italy, the

159 Netherlands, Norway, Spain, Sweden, and the United Kingdom) (27). A total of 521,324 participants

160 were recruited into EPIC between 1992 and 2000. Detailed data on lifestyle, dietary and socio-

161 demographic factors were collected at baseline from all the participants. Body weight and standing

162   height were measured by trained health professionals using standardized protocols. Lifestyle variables

163   such as smoking, physical activity, and the level of education were collected using a validated,

164   standardized questionnaire. Information on the highest attained educational level was categorized as

165   none, primary, technical and professional, secondary, or higher (college or university). Smoking was

166   collected as status (current, past, never), by type of products (cigarettes, cigars, pipe), intensity

167   (number of cigarettes), and duration (in years of smoking). Information on past smoking habits and

168   the years since quitting smoking was collected in former smokers.Physical activity was defined

169   according to the Cambridge physical activity definitions: inactive (sedentary job plus no recreational

170   activity), moderately inactive (sedentary job with < 0.5 h recreational activity daily/or standing job

171   with no recreational activity), moderately active (sedentary job with 0.5 to 1 h recreational activity

172   daily/ or standing job with 0.5 h recreational activity daily/ or physical job with no recreational

173   activity) or active (sedentary job with >1 h recreational activity daily/or standing job with >0.5 h

174   recreational activity daily/or physical job with at least some recreational activity/or heavy manual job)

175   (28). Blood samples were collected and are stored in liquid nitrogen (−196°C) in biobank facilities

176   located at the International Agency for Research on Cancer (IARC), or in local biobanks in Denmark

177   (-150°C) and Sweden (-80°C at Malmö and Umeå) until analysis. Informed consent was obtained

178   from all the participants. The EPIC study was approved by the IARC Ethical Committee and the local

179   ethics committees pertaining to each participating centre.

180

181   Follow-up for CRC incidence and vital status

182   Vital status (98.4% complete) was ascertained on a regular basis using record linkage with centralised

183   regional cancer registries (Denmark, Italy, the Netherlands, Norway, Spain, Sweden, and UK) or via a

184   combination of methods including use of health insurance records, connection with cancer and

185   pathology registries, and active follow-up through participants and their close relatives (France,

186   Germany, and Greece). Incident CRC cases were ascertained according to the classification by the

187   International Classification of Diseases for Oncology (ICD-O, codes C18-C20). Colon cancer

188   included tumours in the proximal site (C18.0-C18.5: from cecum to splenic flexure) or the distal

189   segment (C18.6-C18.7: from descending colon down to sigmoid colon), while rectal cancer included

190   tumours that occurred from the recto-sigmoid junction (C19) down to the rectum (C20). Tumours that

191   arose in the anus and in the anal canal (C21) were not included in this analysis.

192

193   Nested case-control design study

194   A total of 1,416 incident CRC cases were identified and matched on 1:1 ratio to controls by incidence

195   density sampling from all cohort members alive and free of cancer at the time of diagnosis of the

196   index case. Cases were selected sequentially in the order of date of diagnosis and based on sufficient

197   biological sample availability. The following matching criteria were applied: age at blood collection

198   ($\pm$1 year), sex, recruiting centre, time of the day at blood collection ($\pm$3 hours), fasting status at blood

199   collection (<3, 3-6, and >6 hours); and additionally, among women by menopausal status (pre-

200   menopause, perimenopause, and post-menopause), and hormone replacement therapy (HRT) use at

201   time of blood collection (yes/no). We excluded subjects within incomplete matched case sets (i.e. a

202   case without a control or vice versa, n=12), and 26 cases and their matched controls from Greece due

203   to unforeseen data restriction issues. Thus, the final data analysis included 1,378 CRC cases and their

204   matched controls.

205

206   Laboratory analyses of AGEs

207   Plasma concentrations of protein-bound AGEs were determined with Ultra Performance Liquid

208   Chromatography tandem Mass Spectrometry (UPLC-MS/MS) as previously described (29,30). In

209   brief, protein-bound CML, CEL and MG-H1 were extracted from plasma using butanolic

210   hydrochloric acid. The individual AGEs were quantified by calculating the area ratio of each

211   unlabelled peak area to the corresponding internal standard. The sum of AGEs (ΣAGEs, in nmol/L)

212   was calculated by summing up the circulating concentrations of CML, CEL and MG-H1 for each

213   subject. We further calculated the ratios of the AGEs considering their dicarbonyl intermediates:

214   MGO-derived:GO derived (i.e. CEL+MG-H1 divided by CML) (**Figure 1**). We also calculated the

215   ratio of [CEL:MG-H1] to assess the influence of the relative abundance of lysine-sourced MGO-

216   derived AGEs (CEL) vs. arginine-sourced MGO-derived AGEs (MG-H1).

217

218   Statistical analysis

219   Means, standard deviations or frequencies were calculated for all variables. Multivariable conditional

220   logistic regression was used to estimate ORs and 95% CIs for CRC risk associated with circulating

221   levels of protein-bound CML, CEL, MG-H1, ΣAGEs, as well as [CEL+MG-H1):CML] (i.e MGO:GO

222   AGEs), and [CEL:MG-H1]. For each main outcome variable (measured biomarker or calculated

223   ratio), quintile cut-points were determined based on the distribution in controls. We ran two models:

224   model 1 was conditioned on the matching factors; model 2 was further adjusted for body mass index

225   (BMI, continuous), height (continuous), highest attained education level (none, primary, technical and

226   professional, secondary, higher), physical activity (inactive, moderately inactive, moderately active,

227   active), smoking status/duration/intensity (never; current smokers 1-<=15, 16-<=25, >26

228   cigarettes/day; former smokers <=10, 11-<=20, >20 years, occasional), and baseline intake levels of

229   energy (continuous, kcal/day), alcohol, red and processed meats, dietary fibre, and dairy products (all

230   as continuous variables and as g/day). Tests for trend were run by using the median value of each

231   quintile included in the model as continuous variables. Separate sub-group analyses were run by sex

232   and anatomical sub-sites of CRC site (colon, rectal). The heterogeneity of the associations by sex,

233   across anatomical sub-sites and in various sub-groups was assessed using the likelihood ratio test. We

234   assessed the AGEs-CRC association by sub-groups of type-2 diabetes (yes/no; self-reported at

235   baseline) and obesity (defined as BMI ≥ 30 kg/m^2). The potential bias of reverse causality in the

236   AGEs-CRC association was assessed by excluding cases diagnosed within the first two years. All the

237   analyses were conducted using Stata 14.0 (StataCorp, College Station, TX, USA). Two-sided *P*-

238   values <0.05 were statistically significant.

239

240   **Results**

241   Selected baseline characteristics of the study participants are shown in **Table 1**. Compared to controls,

242   cases had higher BMI, and higher intakes of alcohol and red and processed meats, and lower intakes

243   of fruits, vegetables, and dairy products. In addition, cases tended to be less physically active

244   compared to controls.

9

245   The associations of individual and combined AGEs, and their various ratios calculated based

246   on pathway of AGEs derivation are shown in **Table 2**. No statistically significant association was

247   observed between CEL and CRC risk (OR comparing highest to lowest quintile, $OR_{Q5vs.Q1}$=0.88,

248   95%CI:0.64-1.19, P_{trend}=*0.580*), whereas inverse associations for CRC were observed for both CML

249   ($OR_{Q5vs.Q1}$=0.40, 95%CI:0.27-0.59, P_{trend}<*0.001*) and MG-H1 ($OR_{Q5vs.Q1}$=0.73, 95%CI:0.53-1.00,

250   P_{trend}=*0.016*). A near 50% lower odds for developing CRC was observed for ΣAGEs ($OR_{Q5vs.Q1}$=0.52,

251   95%CI:0.37-0.73, P_{trend}<*0.001*), mostly driven by CML and MG-H1. The ratio of [(CEL+MG-H1):

252   CML] was associated with an increased risk or CRC ($OR_{Q5vs.Q1}$=1.91, 95%CI:1.31 - 2.79,

253   P_{trend}=*0.004*).

254   We did not observe significant heterogeneity by sex or by tumour anatomical subsites for

255   individual AGEs, ΣAGEs, and [CEL+MG-H1):CML] **(Table 3).** Analyses stratified by baseline

256   diabetes status and by obesity as indicated by BMI≥ 30 kg/m^2 showed that [(CEL+MG-H1):CML]

257   was associated with higher CRC risk in diabetic vs. non-diabetic subjects, and in obese vs. non-obese

258   subjects (**Table 4**). The inverse associations observed with individual AGEs and for ΣAGEs were

259   more prominent in obese individuals compared to non-obese ones.

260

261   **Discussion**

262   In this study, we found that higher circulating levels of protein-bound CML, MG-H1 and ΣAGEs, but

263   not CEL, were associated with a lower risk of CRC. We also observed that higher concentrations of

264   MGO-derived AGEs relative to GO-derived AGEs were associated with higher CRC risk.

265   The inverse associations observed between the AGEs concentrations evaluated in our study

266   and CRC risk contrast with our hypothesis that these AGEs contribute to colorectal carcinogenesis.

267   Specifically, for CML it is noteworthy that Jiao et al. (22) also reported similar inverse associations as

268   observed in our study. However, their study was based on a sub-population of male, Finnish smokers

269   and applied an ELISA-based methodology to assess relative differences of CML between cancer cases

270   and controls. Unfortunately, ELISA methods have limited reproducibility and do not differentiate

271   between protein-bound and free fractions of circulating AGEs. Our initial hypothesis for positive

272   AGEs-CRC risk association was based on evidence from experimental studies indicating that AGEs

273　are DNA-damaging and can directly induce sustained inflammation in colon tissues through binding

274　with RAGE (17,19,31). We chose to analyze protein-bound AGEs because these are specifically

275　recognized by RAGE (32-34).

276　　　　We assessed CML, CEL and MG-H1 because they are thought to be the most abundant AGEs

277　in the body and although there is evidence that they have harmful pro-inflammatory and pro-oxidative

278　effects, their relationship with the colonic mucosa may be very complicated. The colonic mucosa may

279　be exposed to a variety of other AGEs, both exogenously from the diet and those produced

280　endogenously within the body and colonic milieu. Similar to endogenous AGEs, dietary AGEs may

281　increase the body AGEs pool, interact directly with the colon mucosal and increase the risk for

282　obesity (35). Some researchers have classified AGEs into two categories of "non-toxic" and "toxic"

283　(36-38). This categorization still requires much further evidence, but it is noteworthy that we have

284　previously shown an increased risk of rectal cancer with higher circulating levels of glyceraldehyde-

285　derived AGEs (39), that have been categorized as "toxic" (6,36). Another speculative explanation of

286　our observations is that circulating AGEs concentration may not be reflective of their levels in

287　colorectal tissues where they may accumulate. There is a paucity of supportive data for this assertion,

288　and little is known about whether the concentration and actions of specific AGEs may differ between

289　tissues. Van Heijst et al. (40) observed varying AGEs levels from an immunohistochemical

290　expression of CML and the MGO-derived AGE argpyrimidine in various human tumours (muscle,

291　colon, breast and larynx) suggesting that separate AGEs impact tissues differently. Therefore, relevant

292　studies focusing on the functions, effects, and the interactions of CML, CEL and MG-H1 and

293　additional AGEs within normal and cancerous colonic tissues are warranted.

294　　　　The positive association observed with the ratio of MGO-derived AGEs:GO-derived AGEs

295　and CRC risk suggests that the conditions that may lead to more MGO-AGEs vs. GO-AGEs may be

296　important in CRC development. This result may implicate that higher circulating MGO versus GO

297　may be of greater importance in CRC development than AGEs. Both MGO and GO are mainly

298　detoxified through the glyoxalase (GLO) system and by other enzymes such as aldo-keto reductases

299　and dicarbonyl and L-xylulose reductase (41,42). Compared to GO, MGO is more reactive (43) but is

300　rapidly and efficiently detoxified, mainly in the liver (42). This may possibly explain the higher CRC

11

301 risk observed with MGO-AGE-GO-AGEs in obese individuals. Obesity is often associated with a

302 degree of liver steatosis and decreased liver function and could possibly explain lower clearance of

303 MGO with spillover into the circulation. It can be speculated that in the presence of a "competition"

304 between the production and the detoxification of dicarbonyls, MGO and its derived AGEs may be

305 harmful to the colon tissue, and sustain systemic inflammation, compared to GO and derived AGEs –

306 but this requires further investigation. Diabetes has been associated with a higher risk of CRC (44).

307 Because diabetes is associated with poor glycaemic control, hyperglycaemia, and enhanced

308 production of AGEs, one would expect that the AGEs-CRC association is higher in subjects with

309 diabetes compared to those without. Additional studies should explore whether CRC risk associated

310 with diabetes could be partially mediated through AGEs. Likewise, future studies may also explore to

311 which extent treatment for diabetes may mitigate endogenous AGEs production and possibly CRC

312 risk.

313   It is also noteworthy that dicarbonyls and some AGEs derived from them display hormetic

314 properties, where lower levels are associated with beneficial health outcomes while higher levels are

315 deleterious. Hormetic effects have been reported for lower levels of MGO which have been showed to

316 prevent tumour growth, whereas higher levels promote tumour expansion (45). Surprisingly, it has

317 been reported that another MGO-derived AGE, MG-H3, has anti-oxidative properties comparable

318 with those of ascorbic acid (46). There is substantial evidence showing that the deleterious effects of

319 AGEs are dependent upon the level of RAGE activity. In RAGE knockout mice, cancer development

320 is greatly reduced, suggesting that the cancer-promotive and pro-inflammatory effects of AGEs are

321 necessarily expressed in the presence of RAGE (47,48). Interestingly, there is mounting evidence

322 showing that soluble RAGE (sRAGE), a free circulating isoform of RAGE, is inversely associated

323 with CRC (21,49). The knowledge of AGEs metabolism and CRC need to be expanded, and

324 additional studies are needed to better understand the role of dicarbonyls, and derived AGEs in the

325 aetiology of CRC.

326   This study has several strengths, including the quantitative measurement of CEL, CML and

327 MG-H1 by a state-of-the-art UPLC-MS/MS instrumental method. UPLC-MS/MS could be considered

328 the gold standard method for the analysis of AGEs in plasma. UPLC-MS/MS could be used to

329   accurately and precisely measure specific AGEs in both free and protein-bound forms. The major

330   known drawback of using UPLC-MS/MS is its relatively higher cost and the necessity for trained

331   personnel (23,50). Additional strengths include the prospective design, the large sample size, and the

332   ability to conduct analyses stratified by sex, and by anatomical subsite (colon vs rectum). A limitation

333   to our study is that we lack information on other AGEs produced from MGO including other MGO-

334   derived hydroimidazolone (MG-Hs) such as MG-H2, MG-H3 and MG-H4, Nd-(4-carboxy-4,6-

335   dimethyl5,6-dihydroxy-1,4,5,6-tetrahydropyrimidin-2-yl)-ornithine (THP), argpyrimidine and

336   crosslinking dimer MOLD (8), which may all have roles in CRC development. Another limitation is

337   the use of plasma AGEs levels which are dependent on kidney and liver functions; hence they may

338   not represent tissue levels. Further research is required to determine how circulating AGEs measures

339   in the same individual may relate to levels in colon tumour and normal colon tissues. Therefore, our

340   assessment of AGEs in CRC development is far from complete, even though we analysed three major

341   AGEs compounds. Also, our findings show that circulating measures of AGEs are likely to have

342   differential associations with CRC, indicating that their posited detrimental properties may not be

343   equivalent or that they vary in their pro-inflammatory capacity. More study is required on the

344   individual and interactive roles of AGEs in the development of cancers and other chronic diseases. A

345   deeper assessment of the qualitative pathways of AGEs production and their cumulative roles in

346   cancer development may shed more insight into this fascinating topic. An additional limitation of this

347   nested case-control study is the fact that blood samples and lifestyle factors were collected at baseline

348   and may not necessarily reflect changes over time.

349   In conclusion, in this large, comprehensive prospective study CML and MG-H1 are inversely

350   associated with CRC risk, contrary to our initial hypothesis. However, we observed a significantly

351   higher CRC risk with higher ratio of MGO-derived:GO-dervied AGEs. Our observations highlight the

352   complexity of the proposed roles of AGEs in CRC development and suggest that AGEs levels may

353   not be interpreted alone, but in consideration of their chemical origins. Additional studies examining

354   toxic dicarbonyl AGEs precursor compounds in CRC development, and assessing the role of AGEs in

355   the colonic milieu and within normal and tumorigenic colonic tissues are required. In addition, the

356   development of laboratory instrumental methodologies for the assessment of a larger number of AGEs

357   would aid greatly in better defining the roles of this diverse family of compounds in health and

358   disease.

359

References

1. Bray, F*., et al.* (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*.

2. Safiri, S*., et al.* (2019) The global, regional, and national burden of colorectal cancer and its attributable risk factors in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Gastroenterology & Hepatology*, **4**, 913-933.

3. Murphy, N*., et al.* (2019) Lifestyle and dietary environmental factors in colorectal cancer susceptibility. *Mol Aspects Med*.

4. Mehta, R.S*., et al.* (2017) Dietary Patterns and Risk of Colorectal Cancer: Analysis by Tumor Location and Molecular Subtypes. *Gastroenterology*, **152**, 1944-1953.e1.

5. Cerami, C*., et al.* (1997) Tobacco smoke is a source of toxic reactive glycation products. *Proc Natl Acad Sci U S A*, **94**, 13915-20.

6. Takeuchi, M*., et al.* (2000) Immunological evidence that non-carboxymethyllysine advanced glycation end-products are produced from short chain sugars and dicarbonyl compounds in vivo. *Mol Med*, **6**, 114-25.

7. Scheijen, J*., et al.* (2016) Analysis of advanced glycation endproducts in selected food items by ultra-performance liquid chromatography tandem mass spectrometry: Presentation of a dietary AGE database. *Food Chem*, **190**, 1145-1150.

8. Sharma, C*., et al.* (2015) Advanced glycation End-products (AGEs): an emerging concern for processed food industries. *J Food Sci Technol*, **52**, 7561-76.

9. Schalkwijk, C.G*., et al.* (2020) Methylglyoxal, a Highly Reactive Dicarbonyl Compound, in Diabetes, Its Vascular Complications, and Other Age-Related Diseases. *Physiological Reviews*, **100**, 407-461.

10. Turk, Z. (2010) Glycotoxines, carbonyl stress and relevance to diabetes and its complications. *Physiol Res*, **59**, 147-56.

11. Thornalley, P.J. (2005) Dicarbonyl intermediates in the maillard reaction. *Ann N Y Acad Sci*, **1043**, 111-7.

12. Nowotny, K*., et al.* (2018) Dietary advanced glycation end products and their relevance for human health. *Ageing Res Rev*, **47**, 55-66.

13. Arribas-Lorenzo, G*., et al.* (2010) Analysis, Distribution, and Dietary Exposure of Glyoxal and Methylglyoxal in Cookies and Their Relationship with Other Heat-Induced Contaminants. *Journal of Agricultural and Food Chemistry*, **58**, 2966-2972.

14. Raupbach, J*., et al.* (2020) Proteasomal degradation of glycated proteins depends on substrate unfolding: Preferred degradation of moderately modified myoglobin. *Free Radical Biology and Medicine*, **152**, 516-524.

15. Palanissami, G*., et al.* (2018) RAGE and Its Ligands: Molecular Interplay Between Glycation, Inflammation, and Hallmarks of Cancer-a Review. *Horm Cancer*, **9**, 295-325.

16. Sakellariou, S*., et al.* (2016) Clinical significance of AGE-RAGE axis in colorectal cancer: associations with glyoxalase-I, adiponectin receptor expression and prognosis. *BMC Cancer*, **16**, 174.

17. Kuniyasu, H*., et al.* (2003) Differential effects between amphoterin and advanced glycation end products on colon cancer cells. *Int J Cancer*, **104**, 722-7.

18. Chen, H*., et al.* (2014) Advanced glycation end products increase carbohydrate responsive element binding protein expression and promote cancer cell proliferation. *Mol Cell Endocrinol*, **395**, 69-78.

19. Shimomoto, T*., et al.* (2012) Advanced glycation end products (AGE) induce the receptor for AGE in the colonic mucosa of azoxymethane-injected Fischer 344 rats fed with a high-linoleic acid and high-glucose diet. *J Gastroenterol*, **47**, 1073-83.

20. Wang, P*., et al.* (2018) Advanced Glycation End Products Increase MDM2 Expression via Transcription Factor KLF5. *J Diabetes Res*, **2018**, 3274084.

447   21.   Chen, L.*, et al.* (2016) A prospective study of soluble receptor for advanced glycation end-
448   products and colorectal cancer risk in postmenopausal women. *Cancer Epidemiol*, **42**, 115-
449   23.
450   22.   Jiao, L.*, et al.* (2011) Advanced glycation end products, soluble receptor for advanced
451   glycation end products, and risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev*, **20**,
452   1430-8.
453   23.   de Vos, L.C.*, et al.* (2016) Advanced glycation end products: An emerging biomarker for
454   adverse outcome in patients with peripheral artery disease. *Atherosclerosis*, **254**, 291-299.
455   24.   Baldensperger, T.*, et al.* (2020) Comprehensive analysis of posttranslational protein
456   modifications in aging of subcellular compartments. *Scientific Reports*, **10**, 7596.
457   25.   Ahmed, M.U.*, et al.* (1997) N-epsilon-(carboxyethyl)lysine, a product of the chemical
458   modification of proteins by methylglyoxal, increases with age in human lens proteins.
459   *Biochem J*, **324 (Pt 2)**, 565-70.
460   26.   Verzijl, N.*, et al.* (2000) Age-related accumulation of Maillard reaction products in human
461   articular cartilage collagen. *Biochem J*, **350 Pt 2**, 381-7.
462   27.   Riboli, E.*, et al.* (2002) European Prospective Investigation into Cancer and Nutrition (EPIC):
463   study populations and data collection. *Public Health Nutr*, **5**, 1113-24.
464   28.   Wareham, N.J.*, et al.* (2002) Validity and repeatability of the EPIC-Norfolk Physical Activity
465   Questionnaire. *Int J Epidemiol*, **31**, 168-74.
466   29.   Scheijen, J.*, et al.* (2018) Dietary intake of advanced glycation endproducts is associated with
467   higher levels of advanced glycation endproducts in plasma and urine: The CODAM study. *Clin
468   Nutr*, **37**, 919-925.
469   30.   Maasen, K.*, et al.* (2019) High dietary glycemic load is associated with higher concentrations
470   of urinary advanced glycation endproducts: the Cohort on Diabetes and Atherosclerosis
471   Maastricht (CODAM) Study. *Am J Clin Nutr*, **110**, 358-366.
472   31.   Stopper, H.*, et al.* (2003) Genotoxicity of advanced glycation end products in mammalian
473   cells. *Cancer Lett*, **190**, 151-6.
474   32.   Xue, J.*, et al.* (2014) The receptor for advanced glycation end products (RAGE) specifically
475   recognizes methylglyoxal-derived AGEs. *Biochemistry*, **53**, 3327-35.
476   33.   Xue, J.*, et al.* (2011) Advanced glycation end product recognition by the receptor for AGEs.
477   *Structure (London, England : 1993)*, **19**, 722-732.
478   34.   Xie, J.*, et al.* (2008) Structural basis for pattern recognition by the receptor for advanced
479   glycation end products (RAGE). *J Biol Chem*, **283**, 27255-69.
480   35.   Cordova, R.*, et al.* (2020) Dietary intake of advanced glycation end products (AGEs) and
481   changes in body weight in European adults. *Eur J Nutr*, **59**, 2893-2904.
482   36.   Takata, T.*, et al.* (2019) Intracellular toxic advanced glycation end-products in
483   cardiomyocytes may cause cardiovascular disease. *Scientific Reports*, **9**, 2121.
484   37.   Takeuchi, M.*, et al.* (2000) Neurotoxicity of advanced glycation end-products for cultured
485   cortical neurons. *J Neuropathol Exp Neurol*, **59**, 1094-105.
486   38.   Sato, T.*, et al.* (2006) Toxic advanced glycation end products (TAGE) theory in Alzheimer's
487   disease. *Am J Alzheimers Dis Other Demen*, **21**, 197-208.
488   39.   Kong, S.Y.*, et al.* (2015) The Association between Glyceraldehyde-Derived Advanced
489   Glycation End-Products and Colorectal Cancer Risk. *Cancer Epidemiol Biomarkers Prev*, **24**,
490   1855-63.
491   40.   van Heijst, J.W.*, et al.* (2005) Advanced glycation end products in human cancer tissues:
492   detection of Nepsilon-(carboxymethyl)lysine and argpyrimidine. *Ann N Y Acad Sci*, **1043**, 725-
493   33.
494   41.   Kalousova, M.*, et al.* (2010) Genetic predisposition to advanced glycation end products
495   toxicity is related to prognosis of chronic hemodialysis patients. *Kidney Blood Press Res*, **33**,
496   30-6.

497 42. Yang, K.*, et al.* (2011) Differences in glyoxal and methylglyoxal metabolism determine
498   cellular susceptibility to protein carbonylation and cytotoxicity. *Chem Biol Interact*, **191**, 322-
499   9.
500 43. Maessen, D.E.*, et al.* (2016) Energy restriction and Roux-en-Y gastric bypass reduce
501   postprandial α-dicarbonyl stress in obese women with type 2 diabetes. *Diabetologia*, **59**,
502   2013-7.
503 44. Ma, Y.*, et al.* (2018) Type 2 diabetes and risk of colorectal cancer in two large U.S.
504   prospective cohorts. *Br J Cancer*, **119**, 1436-1442.
505 45. Morgenstern, J.*, et al.* (2020) The Glyoxalase System-New Insights into an Ancient
506   Metabolism. *Antioxidants (Basel)*, **9**.
507 46. Wang, T.*, et al.* (2012) Exploring post-translational arginine modification using chemically
508   synthesized methylglyoxal hydroimidazolones. *J Am Chem Soc*, **134**, 8958-67.
509 47. Kwak, T.*, et al.* (2017) Targeting of RAGE-ligand signaling impairs breast cancer cell invasion
510   and metastasis. *Oncogene*, **36**, 1559-1572.
511 48. Gebhardt, C.*, et al.* (2008) RAGE signaling sustains inflammation and promotes tumor
512   development. *The Journal of experimental medicine*, **205**, 275-285.
513 49. Aglago, E.K.*, et al.* (2020) Soluble Receptor for Advanced Glycation End-products (sRAGE)
514   and Colorectal Cancer Risk: A Case-Control Study Nested within a European Prospective
515   Cohort. *Cancer Epidemiol Biomarkers Prev*.
516 50. Perrone, A.*, et al.* (2020) Advanced Glycation End Products (AGEs): Biochemistry, Signaling,
517   Analytical Methods, and Epigenetic Effects. *Oxid Med Cell Longev*, **2020**, 3818196.

518

519

FIGURES LEGENDS
521 **Figure 1:** Schematic representation of the formation of the AGEs and the rationale for the calculation
522 of the ratios
523 Abbreviations: AGE, advanced glycation end-product; CML, N^ε-carboxy-methyllysine; CEL, N^ε-
524 carboxy-ethyllysine; GO, glyoxal; MG-H1, N^δ-(5-hydro-5-methyl-4-imidazolon-2-yl)-ornithine;
525 MGO, methylglyoxal
526 AGEs are absorbed from the diet or formed during the Maillard reactions from the Amadori or Heyns
527 products and from the glycating actions of dicarbonyls such as MGO and GO. CML is derived from
528 GO whereas MG-H1 and CEL are derived from MGO.
529

Table 1: Selected baseline demographic and lifestyle characteristics of study participants by colorectal cancer status,

	Cases (1,378)	Controls (n=1,378)
Women, %	51.7	51.5
Anthropometry, mean (SD)		
BMI, kg/m²	26.7±4.25	26.2±3.74
Waist circumference, cm	90.4±13.0	88.3±12.1
Waist-to-hip ratio	0.88±0.10	0.87±0.10
Lifestyle variables, n (%)		
Smoking status frequency and intensity		
Never	542 (39.8)	514 (37.9)
Current, 1 - 15 cig/day	139 (10.2)	129 (9.51)
Current, 16-25 cig/day	94 (6.91)	87 (6.40)
Current, 26+ cig/day	23 (1.69)	20 (1.47)
Former, quit <= 10 years	129 (9.48)	139 (10.3)
Former, quit 11-20 years	123 (9.04)	144 (10.6)
Former, quit 20+ years	177 (13.0)	166 (12.2)
Current, pipe/cigar/occasional	102 (7.49)	125 (9.22)
Physical activity		
Inactive	361 (25.9)	327 (23.3)
Moderately inactive	448 (32.1)	457 (32.6)
Moderately active	311 (22.3)	284 (20.3)
Active	263 (18.9)	314 (22.4)
Highest education level attained		
None	66 (4.85)	68 (5.01)
Primary school completed	490 (36.0)	453 (33.4)
Technical/professional school	343 (25.2)	324 (23.9)
Secondary school	184 (13.5)	217 (16.0)
Higher education	244 (17.9)	247 (18.2)
Dietary intake, mean (SD)		
Energy, Kcal/day	2127±609	2124±620
Alcohol, g/day	17.0±22.1	15.4±19.7
Red and processed meats, g/day	87.6±53.1	85.1±52.0
Fruits and vegetables, g/day	396±233	421±248
Cereals, g/day	216±121	216±119
Dairy products, g/day	331±251	351±244
Fish and products, g/day	28.2±28.8	29.6±30.6
Sugar, cakes and confectionaries, g/day	48.7±66.6	48.7±68.9
Fats, g/day	28.3±15.6	27.9±16.0
Protein, g/day	89.3±27.9	90.3±27.5
AGEs Biomarkers, mean (SD)		
CML, nmol/l	2719±1046	2855±107 5
CEL, nmol/l	1475±772	1475±740

MGH1, nmol/l	1056±259	1079±262
ΣAGEs, nmol/l	5250±1488	5411±1470
CEL:MG-H1	1.45±0.79	1.43±0.75
(CEL+MG-H1): CML	1.01±0.39	0.98±0.38

532 Frequencies may not add up to 100% due to missing data
533 Abbreviations: AGE, Advanced glycation end products; BMI, Body mass index; CML, Nε-
534 carboxymethyl-lysine; CEL, Nε-carboxyethyl-lysine; MG-H1, Nδ-(5-hydro-5-methyl-4-imidazolon-
535 2-yl) ornithine
536

Table 2: ORs and 95%CI for colorectal cancer risk associated with circulating AGEs and their ratios, EPIC study 1992-2012

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	P_{trend}
CML						
Range, nmol/L	<2014	2014-<2401	2401-<2805	2805-<3505	≥3505	
Cases/controls	302/276	338/276	280/275	247/276	211/275	
Model 1*	1.00 (Ref.)	0.96 (0.75 - 1.23)	0.74 (0.57 - 0.97)	0.58 (0.44 - 0.78)	0.37 (0.26 - 0.52)	<0.001
Model 2†	1.00 (Ref.)	0.97 (0.75 - 1.26)	0.79 (0.59 - 1.04)	0.63 (0.46 - 0.86)	0.40 (0.27 - 0.59)	<0.001
CEL						
Range, nmol/L	<986	986-<1234	1234-<1478	1478-<1807	≥1807	
Cases/controls	272/276	270/276	286/276	290/275	260/275	
Model 1*	1.00 (Ref.)	0.96 (0.74 - 1.23)	1.02 (0.78 - 1.33)	1.01 (0.76 - 1.33)	0.89 (0.66 - 1.19)	0.605
Model 2†	1.00 (Ref.)	0.98 (0.76 - 1.27)	1.04 (0.79 - 1.37)	1.04 (0.78 - 1.39)	0.88 (0.64 - 1.19)	0.580
MG-H1						
Range, nmol/L	<872	872-<974	974-<1082	1082-<1248	≥1248	
Cases/controls	309/276	308/275	260/276	256/274	244/276	
Model 1*	1.00 (Ref.)	0.94 (0.73 - 1.20)	0.75 (0.58 - 0.97)	0.71 (0.54 - 0.93)	0.68 (0.50 - 0.91)	0.002
Model 2†	1.00 (Ref.)	0.97 (0.75 - 1.25)	0.79 (0.61 - 1.03)	0.77 (0.58 - 1.02)	0.73 (0.53 - 1.00)	0.016
ΣAGEs, nmol/L						
Range, nmol/L	<4284	4284-<4848	4848-<5414	5414-<6306	≥6306	
Cases/controls	334/276	315/276	275/276	219/276	235/274	
Model 1*	1.00 (Ref.)	0.89 (0.70 - 1.14)	0.73 (0.56 - 0.95)	0.52 (0.39 - 0.68)	0.48 (0.35 - 0.65)	<0.001
Model 2†	1.00 (Ref.)	0.93 (0.72 - 1.19)	0.76 (0.58 - 1.00)	0.54 (0.41 - 0.73)	0.52 (0.37 - 0.73)	<0.001
CEL:MG-H1						
Range	<0.89	0.89-<1.15	1.15-<1.43	1.43-<1.81	≥1.81	
Cases/controls	247/276	274/276	263/275	295/275	298/275	
Model 1*	1.00 (Ref.)	1.09 (0.85 - 1.40)	1.09 (0.83 - 1.42)	1.27 (0.96 - 1.68)	1.33 (0.98 - 1.80)	0.047
Model 2†	1.00 (Ref.)	1.13 (0.87 - 1.47)	1.08 (0.82 - 1.42)	1.27 (0.95 - 1.70)	1.26 (0.91 - 1.73)	0.139
(CEL+MG-H1): CML						
Range	<0.66	0.66-<0.86	0.86-<1.02	1.02-<1.24	≥1.24	
Cases/controls	233/276	279/276	263/275	280/275	322/275	

	Q1	Q2	Q3	Q4	Q5	P-trend
Model 1*	1.00 (Ref.)	1.49 (1.12 - 1.99)	1.64 (1.19 - 2.27)	1.70 (1.21 - 2.39)	2.14 (1.50 - 3.05)	*<0.001*
Model 2†	1.00 (Ref.)	1.42 (1.05 - 1.90)	1.54 (1.10 - 2.16)	1.54 (1.08 - 2.19)	1.91 (1.31 - 2.79)	*0.004*

Abbreviations: AGE, advanced glycation end-product; CI, confidence interval; CML, N^{ε}-carboxy-methyllysine; CEL, N^{ε}-carboxy-ethyllysine; MG-H1, N^{δ}-(5-hydro-5-methyl-4-imidazolon-2-yl)-ornithine; OR, odds ratio; Quintiles were created based on the distribution in the control group

MG-H1 has one missing data, hence MG-H1, CEL:MG-H1 and (CEL+MG-H1): CML have 1,377 cases and 1,377 matched controls

*Model 1 was conditioned on matching factors: age at blood collection (±1 year), sex, recruiting centre, time of the day at blood collection (±3 hours), fasting status at blood collection (<3, 3-6, and >6 hours); and additionally, among women by menopausal status (pre-menopause, perimenopause, and post-menopause), and hormone replacement therapy (HRT) use at time of blood collection (yes/no)

†Model 2 model was Model 1 adjusted for BMI (continuous), height (continuous), education (none, primary, technical and professional, secondary, higher education), physical activity (inactive, moderately inactive, moderately active, active), smoking status, duration and intensity (never, 1 - 15 cigarettes/day, 16-25 cigarettes/day, over 26 cigarettes/day, former smokers who quit<10 years, former smokers who quit 11-20 years, former smokers who quit>20 years, current pipe-cigar and occasional smokers), energy intake (continuous), alcohol intake (continuous), processed meat intake (continuous), fibre intake (continuous) and dairy products intake (continuous)

Table 3: ORs and 95%CI for colorectal cancer risk associated with one standard deviation increase in circulating AGEs and their ratios, by sex and by tumour anatomical subsite, EPIC study 1992-2012

	Colorectal cancer		Colon cancer		Rectal cancer		$P_{heterogeneity}$ by tumour subsite
	Cases/Controls	OR (95% CI)	Cases/Controls	OR (95% CI)	Cases/Controls	OR (95% CI)	
CML, nmol/L							
All	1378/1378	0.75 (0.66 - 0.85)	871/871	0.69 (0.58 - 0.83)	503/503	0.81 (0.66 - 1.00)	0.073
Men	679/679	0.69 (0.57 - 0.83)	404/404	0.66 (0.51 - 0.86)	272/272	0.67 (0.49 - 0.92)	0.162
Women	699/699	0.81 (0.67 - 0.97)	467/467	0.75 (0.59 - 0.96)	231/231	0.90 (0.65 - 1.26)	0.323
$P_{heterogeneity}$ by sex		0.197		0.223		0.622	
CEL, nmol/L							
All	1378/1378	0.98 (0.88 - 1.08)	871/871	0.98 (0.86 - 1.11)	503/503	1.00 (0.83 - 1.20)	0.986
Men	679/679	1.01 (0.85 - 1.20)	404/404	0.99 (0.77 - 1.25)	272/272	1.10 (0.82 - 1.46)	0.291
Women	699/699	0.97 (0.85 - 1.11)	467/467	1.00 (0.85 - 1.17)	231/231	0.83 (0.61 - 1.12)	0.298
$P_{heterogeneity}$ by sex		0.839		0.596		0.129	
MG-H1, nmol/L							
All	1377/1377	0.88 (0.79 - 0.98)	871/871	0.81 (0.71 - 0.93)	503/503	0.99 (0.83 - 1.17)	0.056
Men	678/678	0.83 (0.71 - 0.97)	404/404	0.77 (0.62 - 0.95)	272/272	0.87 (0.68 - 1.13)	0.173
Women	699/699	0.92 (0.80 - 1.07)	467/467	0.87 (0.72 - 1.05)	231/231	1.01 (0.78 - 1.31)	0.346
$P_{heterogeneity}$ by sex		0.313		0.332		0.701	
ΣAGEs, nmol/L							
All	1377/1377	0.81 (0.72 - 0.91)	871/871	0.76 (0.65 - 0.89)	503/503	0.76 (0.65 - 0.89)	0.144
Men	678/678	0.76 (0.64 - 0.91)	404/404	0.71 (0.55 - 0.91)	272/272	0.81 (0.61 - 1.07)	0.066
Women	699/699	0.85 (0.73 - 1.01)	467/467	0.84 (0.68 - 1.03)	231/231	0.85 (0.62 - 1.15)	0.913
$P_{heterogeneity}$ by sex		0.233		0.083		0.747	
CEL:MG-H1							
All	1377/1377	1.03 (0.93 - 1.14)	871/871	1.05 (0.93 - 1.19)	503/503	0.99 (0.82 - 1.21)	0.517
Men	678/678	1.10 (0.91 - 1.32)	404/404	1.07 (0.83 - 1.37)	272/272	1.22 (0.87 - 1.70)	0.572
Women	699/699	1.03 (0.90 - 1.17)	467/467	1.08 (0.93 - 1.26)	231/231	0.81 (0.60 - 1.09)	0.175
$P_{heterogeneity}$ for sex		0.528		0.967		0.110	

(CEL+MG-H1): CML							
All	1377/1377	1.15 (1.03 - 1.29)	871/871	1.15 (1.00 - 1.32)	503/503	1.17 (0.94 - 1.46)	*0.630*
Men	678/678	1.34 (1.10 - 1.64)	404/404	1.27 (0.99 - 1.64)	272/272	1.65 (1.12 - 2.44)	*0.852*
Women	699/699	1.10 (0.95 - 1.27)	467/467	1.15 (0.97 - 1.37)	231/231	0.89 (0.65 - 1.23)	*0.292*
$P_{heterogeneity}$ by sex		*0.084*		*0.292*		*0.065*	

Abbreviations: AGE, advanced glycation end-product; CI, confidence interval; CML, N^{ε}-carboxy-methyllysine; CEL, N^{ε}-carboxy-ethyllysine; MG-H1, N^{δ}-(5-hydro-5-methyl-4-imidazolon-2-yl)-ornithine; OR, odds ratio

MG-H1 has one missing data, hence MG-H1, CEL:MG-H1 and (CEL+MG-H1): CML have 1,377 cases and 1,377 matched controls

4 cases of overlapping tumours were considered as colorectal cancer cases, but not classified as colon malignant tumour or rectal one.

*Models were conditioned on matching factors: age at blood collection (±1 year), sex, recruiting centre, time of the day at blood collection (±3 hours), fasting status at blood collection (<3, 3-6, and >6 hours); and additionally, among women by menopausal status (pre-menopause, peri-menopause, and post-menopause), and hormone replacement therapy (HRT) use at time of blood collection (yes/no) and adjusted for BMI (continuous), height (continuous), education (none, primary, technical and professional, secondary, higher education), physical activity (inactive, moderately inactive, moderately active, active), smoking status, duration and intensity (never, 1 - 15 cigarettes/day, 16-25 cigarettes/day, over 26 cigarettes/day, former smokers who quit<10 years, former smokers who quit 11-20 years, former smokers who quit>20 years, current pipe-cigar and occasional smokers), energy intake (continuous), alcohol intake (continuous), processed meat intake (continuous), fibre intake (continuous) and dairy products intake (continuous)

Table 4: ORs and 95%CI for colorectal cancer risk associated with one standard deviation increase in circulating AGEs and their ratios, stratified by obesity and diabetes status, EPIC study 1992-2012

	Diabetes*			Obese[†]		
	Yes (n=61)	No (n=1099)	*P* for heterogeneity	Yes (n=247)	No (n=1131)	*P* for heterogeneity
CML	0.56 (0.25 - 1.25)	0.74 (0.64 - 0.85)	*0.486*	0.30 (0.19 - 0.47)	0.85 (0.74 - 0.97)	*<0.001*
CEL	2.07 (0.89 - 4.82)	0.99 (0.88 - 1.10)	*0.088*	1.01 (0.68 - 1.50)	1.01 (0.90 - 1.12)	*0.999*
MGH1	1.20 (0.55 - 2.61)	0.87 (0.78 - 0.98)	*0.443*	0.46 (0.32 - 0.67)	0.96 (0.86 - 1.07)	*<0.001*
ΣAGEs	1.07 (0.48 - 2.40)	0.81 (0.72 - 0.92)	*0.499*	0.38 (0.25 - 0.57)	0.90 (0.80 - 1.02)	*<0.001*
CEL:MG-H1	1.73 (0.78 - 3.84)	1.05 (0.94 - 1.17)	*0.228*	1.59 (1.11 - 2.27)	1.02 (0.92 - 1.14)	*0.019*
(CEL+MG-H1): CML	2.94 (1.21 - 7.15)	1.19 (1.05 - 1.35)	*0.046*	2.36 (1.60 - 3.47)	1.10 (0.98 - 1.24)	*<0.001*

Abbreviations: AGE, advanced glycation end-product; CI, confidence interval; CML, N^ε-carboxy-methyllysine; CEL, N^ε-carboxy-ethyllysine; MG-H1, N^δ-(5-hydro-5-methyl-4-imidazolon-2-yl)-ornithine; OR, odds ratio

*Self-reported history of diabetes at baseline (n missing=218)

[†]Obesity was defined depending on BMI. Obese: BMI>=30 kg/m², non-obese: BMI<30 kg/m²

*Models were conditioned on matching factors: age at blood collection (±1 year), sex, recruiting centre, time of the day at blood collection (±3 hours), fasting status at blood collection (<3, 3-6, and >6 hours); and additionally, among women by menopausal status (pre-menopause, peri-menopause, and post-menopause), and hormone replacement therapy (HRT) use at time of blood collection (yes/no) and adjusted for height (continuous), education (none, primary, technical and professional, secondary, higher education), physical activity (inactive, moderately inactive, moderately active, active), smoking status, duration and intensity (never, 1 - 15 cigarettes/day, 16-25 cigarettes/day, over 26 cigarettes/day, former smokers who quit<10 years, former smokers who quit 11-20 years, former smokers who quit>20 years, current pipe-cigar and occasional smokers), energy intake (continuous), alcohol intake (continuous), processed meat intake (continuous), fibre intake (continuous) and dairy products intake (continuous)
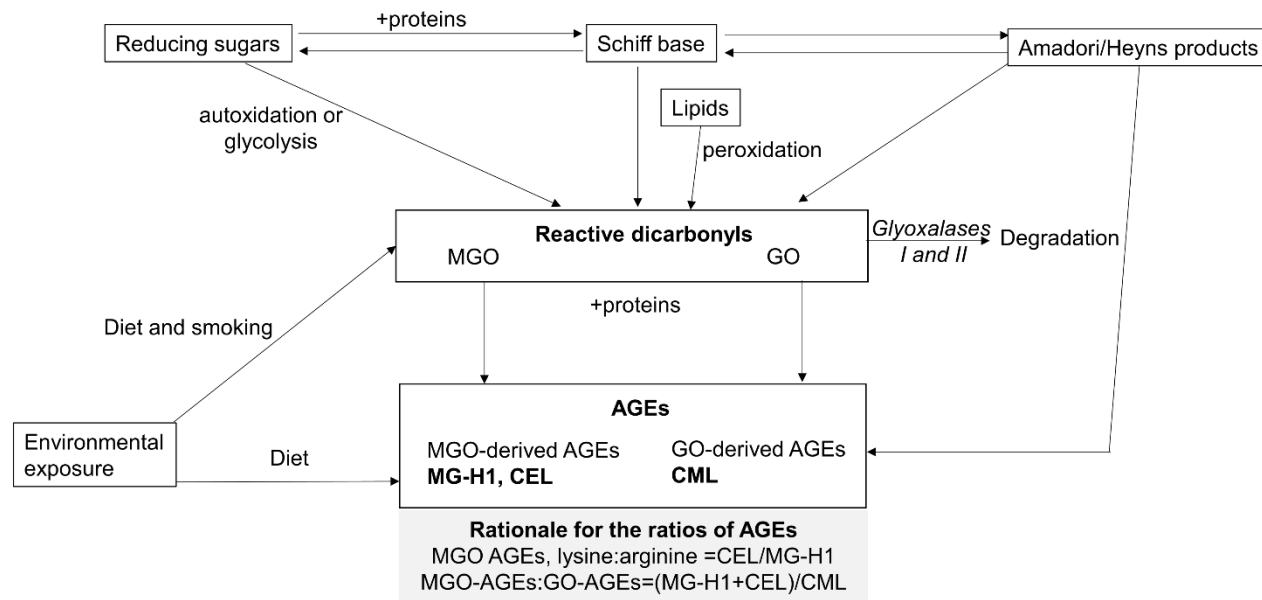
Figure 1