

Plasma concentrations of advanced glycation end-products and colorectal cancer risk in the EPIC study

Short title: AGEs and colorectal cancer

Elom K. Aglago¹, Casper G. Schalkwijk², Heinz Freisling¹, Veronika Fedirko³, David J. Hughes⁴, Li Jiao⁵, Christina C Dahm⁶, Anja Olsen^{6,7}, Anne Tjønneland^{7,8}, Verena Katzke⁹, Theron Johnson⁹, Matthias B. Schulze^{10,11}, Krasimira Aleksandrova^{11,12}, Giovanna Masala¹³, Sabina Sieri¹⁴, Vittorio Simeon¹⁵, Rosario Tumino¹⁶, Alessandra Macciotta¹⁷, Bas Bueno-de-Mesquita¹⁸, Guri Skeie¹⁹, Inger Torhild Gram¹⁹, Torkjel Sandanger¹⁹, Paula Jakszyn^{20,21}, Maria-Jose Sánchez^{22, 23, 24, 25}, Pilar Amiano^{25,26}, Sandra M. Colorado-Yohar^{25,27,28}, Aurelio Barricarte Gurrea^{25, 29, 30}, Aurora Perez-Cornago³¹, Ana-Lucia Mayén¹, Elisabete Weiderpass³², Marc J. Gunter¹, Alicia K. Heath³³, Mazda Jenab¹

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

²Department of Internal Medicine, CARIM School for Cardiovascular Diseases, Maastricht University Medical Center, The Netherlands

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

⁴Cancer Biology and Therapeutics Group (CBT), Conway Institute, School of Biomolecular and Biomedical Science (SBBS), University College Dublin, Ireland

⁵Department of Medicine, Baylor College of Medicine, Houston, TX, USA

⁶Department of Public Health, Aarhus University, Denmark

⁷Danish Cancer Society Research Center, Denmark

⁸Department of Public Health, University of Copenhagen, Denmark

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

¹⁰Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany

¹¹Institute of Nutrition Science, University of Potsdam, Nuthetal, Germany

¹²Nutrition, Immunity and Metabolism Department of Nutrition and Gerontology, German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), Nuthetal, Germany

¹³Cancer Risk Factors and Life-Style Epidemiology Unit, Institute for Cancer Research, Prevention and Clinical Network -ISPRO, Florence, Italy

¹⁴Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milano, Italy

¹⁵Dipartimento di Salute Mentale e Fisica e Medicina Preventiva, University 'Luigi Vanvitelli', Napoli, Italy

¹⁶Cancer Registry and Histopathology Department, Provincial Health Authority (ASP 7), Ragusa, Italy

- ¹⁷Department of Clinical and Biological Sciences, University of Turin, Italy
- ¹⁸Former senior scientist, Dept. for Determinants of Chronic Diseases (DCD), National Institute for Public Health and the Environment (RIVM), PO Box 1, 3720 BA Bilthoven, The Netherlands
- ¹⁹Faculty of Health Sciences, Department of Community Medicine, University of Tromsø, The Arctic University of Norway
- ²⁰Unit of Nutrition and Cancer, Cancer Epidemiology Research Programme, Catalan Institute of Oncology, Barcelona, Spain
- ²¹Blanquerna School of Health Sciences, Ramon Llull University, Barcelona, Spain
- ²²Escuela Andaluza de Salud Pública (EASP), Granada, Spain
- ²³Instituto de Investigación Biosanitaria ibs.Granada, Granada, Spain
- ²⁴Department of Preventive Medicine and Public Health, University of Granada, Granada, Spain.
- ²⁵Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
- ²⁶Public Health Division of Gipuzkoa, BioDonostia Research Institute, Donostia-San Sebastian, Spain
- ²⁷Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain.
- ²⁸Research Group on Demography and Health, National Faculty of Public Health, University of Antioquia, Medellín, Colombia
- ²⁹Navarra Public Health Institute, Pamplona, Spain.
- ³⁰Navarra Institute for Health Research (IdiSNA) Pamplona, Spain.
- ³¹Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, UK
- ³²Office of the Director, International Agency for Research on Cancer (IARC), Lyon, France
- ³³Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, UK

*** To whom correspondence should be addressed.:** Mazda Jenab, Address: 150 Cours Albert Thomas, 69372 Lyon Cedex 08, Email: jenabm@iarc.fr, Tel: +33 472 73 80 82

Keywords: Advanced glycation end-product, glycotoxin, colorectal cancer

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

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

Abbreviations used: AGE, Advanced glycation end-product; BMI, body mass index; CEL, N^ε-(carboxyethyl)lysine; CI, confidence interval; CML, N^ε-(carboxymethyl)lysine; CRC, colorectal cancer; EPIC, European Prospective Investigation into Cancer and Nutrition; IARC, International

75 Agency for Research on Cancer; ICD, international classification of diseases; MG-H1, N δ -(5-hydro-
76 5-methyl-4-imidazol-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-imidazolone-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, apportion 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-(carboxyethyl)lysine, CEL; and N^δ-(5-hydro-5-methyl-4-imidazol-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

possible differences by sex or tumour anatomical subsite. Furthermore, they detected AGEs by ELISA kits, which have low specificity and reproducibility (23).

The aim of the present study was to examine the associations between pre-diagnostic protein-bound circulating levels of CML, CEL, and MG-H1 measured using ultra-performance liquid chromatography tandem mass spectrometry (UPLC-MS/MS) and CRC risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. The rationale for the selection of these three AGEs is threefold: they are considered as the most abundant in the body, they have very well characterized chemical structures, and they are derived from specific pathways of formation with CML being mainly derived from GO whereas CEL and MG-H1 are mainly derived from MGO. For CML and CEL, lysine is the amino-acid of the glycation site, whereas it is arginine for MG-H1. At the cellular level, the MGO-lysine adduct CEL is predominantly formed in the cytosol through the glycation of cytosol proteins whereas the MGO-arginine adduct MG-H1 is equally found in cytosol, histone, and mitochondria proteins (24). We hypothesised that protein-bound concentrations of these AGEs would be associated with a higher CRC risk. We also examined CRC risk associated with the ratios of AGEs from specific dicarbonyls similar to the [CEL:MG-H1] ratio assessed in previous studies (25,26), as a potential index of the chemical origin of the AGEs. Although MGO glyicates amino acids to CEL and MG-H1, both these MGO-derived AGEs have different promoting factors as they are produced from lysine and arginine, respectively. Thus, we applied the ratio of [CEL:MG-H1] as a proxy of the potential differential glycating activities of MGO in the body.

Materials and methods

Study population and data collection

We conducted a nested case-control study within the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort, an ongoing multicentre prospective study with participants recruited from 23 centres constellated in 10 European countries (Denmark, France, Germany, Greece, Italy, the Netherlands, Norway, Spain, Sweden, and the United Kingdom) (27). A total of 521,324 participants were recruited into EPIC between 1992 and 2000. Detailed data on lifestyle, dietary and socio-demographic factors were collected at baseline from all the participants. Body weight and standing

height were measured by trained health professionals using standardized protocols. Lifestyle variables such as smoking, physical activity, and the level of education were collected using a validated, standardized questionnaire. Information on the highest attained educational level was categorized as none, primary, technical and professional, secondary, or higher (college or university). Smoking was collected as status (current, past, never), by type of products (cigarettes, cigars, pipe), intensity (number of cigarettes), and duration (in years of smoking). Information on past smoking habits and the years since quitting smoking was collected in former smokers. Physical activity was defined according to the Cambridge physical activity definitions: inactive (sedentary job plus no recreational activity), moderately inactive (sedentary job with < 0.5 h recreational activity daily/or standing job with no recreational activity), moderately active (sedentary job with 0.5 to 1 h recreational activity daily/ or standing job with 0.5 h recreational activity daily/ or physical job with no recreational activity) or active (sedentary job with >1 h recreational activity daily/or standing job with >0.5 h recreational activity daily/or physical job with at least some recreational activity/or heavy manual job) (28). Blood samples were collected and are stored in liquid nitrogen (-196°C) in biobank facilities located at the International Agency for Research on Cancer (IARC), or in local biobanks in Denmark (-150°C) and Sweden (-80°C at Malmö and Umeå) until analysis. Informed consent was obtained from all the participants. The EPIC study was approved by the IARC Ethical Committee and the local ethics committees pertaining to each participating centre.

Follow-up for CRC incidence and vital status

Vital status (98.4% complete) was ascertained on a regular basis using record linkage with centralised regional cancer registries (Denmark, Italy, the Netherlands, Norway, Spain, Sweden, and UK) or via a combination of methods including use of health insurance records, connection with cancer and pathology registries, and active follow-up through participants and their close relatives (France, Germany, and Greece). Incident CRC cases were ascertained according to the classification by the International Classification of Diseases for Oncology (ICD-O, codes C18-C20). Colon cancer included tumours in the proximal site (C18.0-C18.5: from cecum to splenic flexure) or the distal segment (C18.6-C18.7: from descending colon down to sigmoid colon), while rectal cancer included

tumours that occurred from the recto-sigmoid junction (C19) down to the rectum (C20). Tumours that arose in the anus and in the anal canal (C21) were not included in this analysis.

Nested case-control design study

A total of 1,416 incident CRC cases were identified and matched on 1:1 ratio to controls by incidence density sampling from all cohort members alive and free of cancer at the time of diagnosis of the index case. Cases were selected sequentially in the order of date of diagnosis and based on sufficient biological sample availability. The following matching criteria were applied: 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). We excluded subjects within incomplete matched case sets (i.e. a case without a control or vice versa, $n=12$), and 26 cases and their matched controls from Greece due to unforeseen data restriction issues. Thus, the final data analysis included 1,378 CRC cases and their matched controls.

Laboratory analyses of AGEs

Plasma concentrations of protein-bound AGEs were determined with Ultra Performance Liquid Chromatography tandem Mass Spectrometry (UPLC-MS/MS) as previously described (29,30). In brief, protein-bound CML, CEL and MG-H1 were extracted from plasma using butanolic hydrochloric acid. The individual AGEs were quantified by calculating the area ratio of each unlabelled peak area to the corresponding internal standard. The sum of AGEs (Σ AGEs, in nmol/L) was calculated by summing up the circulating concentrations of CML, CEL and MG-H1 for each subject. We further calculated the ratios of the AGEs considering their dicarbonyl intermediates: MGO-derived:GO derived (i.e. CEL+MG-H1 divided by CML) (**Figure 1**). We also calculated the ratio of [CEL:MG-H1] to assess the influence of the relative abundance of lysine-sourced MGO-derived AGEs (CEL) vs. arginine-sourced MGO-derived AGEs (MG-H1).

Statistical analysis

Means, standard deviations or frequencies were calculated for all variables. Multivariable conditional logistic regression was used to estimate ORs and 95% CIs for CRC risk associated with circulating levels of protein-bound CML, CEL, MG-H1, Σ AGEs, as well as [CEL+MG-H1]:CML] (i.e MGO:GO AGEs), and [CEL:MG-H1]. For each main outcome variable (measured biomarker or calculated ratio), quintile cut-points were determined based on the distribution in controls. We ran two models: model 1 was conditioned on the matching factors; model 2 was further adjusted for body mass index (BMI, continuous), height (continuous), highest attained education level (none, primary, technical and professional, secondary, higher), physical activity (inactive, moderately inactive, moderately active, active), smoking status/duration/intensity (never; current smokers 1- \leq 15, 16- \leq 25, $>$ 26 cigarettes/day; former smokers \leq 10, 11- \leq 20, $>$ 20 years, occasional), and baseline intake levels of energy (continuous, kcal/day), alcohol, red and processed meats, dietary fibre, and dairy products (all as continuous variables and as g/day). Tests for trend were run by using the median value of each quintile included in the model as continuous variables. Separate sub-group analyses were run by sex and anatomical sub-sites of CRC site (colon, rectal). The heterogeneity of the associations by sex, across anatomical sub-sites and in various sub-groups was assessed using the likelihood ratio test. We assessed the AGEs-CRC association by sub-groups of type-2 diabetes (yes/no; self-reported at baseline) and obesity (defined as BMI \geq 30 kg/m²). The potential bias of reverse causality in the AGEs-CRC association was assessed by excluding cases diagnosed within the first two years. All the analyses were conducted using Stata 14.0 (StataCorp, College Station, TX, USA). Two-sided *P*-values $<$ 0.05 were statistically significant.

Results

Selected baseline characteristics of the study participants are shown in **Table 1**. Compared to controls, cases had higher BMI, and higher intakes of alcohol and red and processed meats, and lower intakes of fruits, vegetables, and dairy products. In addition, cases tended to be less physically active compared to controls.

The associations of individual and combined AGEs, and their various ratios calculated based on pathway of AGEs derivation are shown in **Table 2**. No statistically significant association was observed between CEL and CRC risk (OR comparing highest to lowest quintile, $OR_{Q5vs.Q1}=0.88$, 95%CI:0.64-1.19, $P_{trend}=0.580$), whereas inverse associations for CRC were observed for both CML ($OR_{Q5vs.Q1}=0.40$, 95%CI:0.27-0.59, $P_{trend}<0.001$) and MG-H1 ($OR_{Q5vs.Q1}=0.73$, 95%CI:0.53-1.00, $P_{trend}=0.016$). A near 50% lower odds for developing CRC was observed for Σ AGEs ($OR_{Q5vs.Q1}=0.52$, 95%CI:0.37-0.73, $P_{trend}<0.001$), mostly driven by CML and MG-H1. The ratio of [(CEL+MG-H1):CML] was associated with an increased risk of CRC ($OR_{Q5vs.Q1}=1.91$, 95%CI:1.31 - 2.79, $P_{trend}=0.004$).

We did not observe significant heterogeneity by sex or by tumour anatomical subsites for individual AGEs, Σ AGEs, and [(CEL+MG-H1):CML] (**Table 3**). Analyses stratified by baseline diabetes status and by obesity as indicated by $BMI \geq 30 \text{ kg/m}^2$ showed that [(CEL+MG-H1):CML] was associated with higher CRC risk in diabetic vs. non-diabetic subjects, and in obese vs. non-obese subjects (**Table 4**). The inverse associations observed with individual AGEs and for Σ AGEs were more prominent in obese individuals compared to non-obese ones.

Discussion

In this study, we found that higher circulating levels of protein-bound CML, MG-H1 and Σ AGEs, but not CEL, were associated with a lower risk of CRC. We also observed that higher concentrations of MGO-derived AGEs relative to GO-derived AGEs were associated with higher CRC risk.

The inverse associations observed between the AGEs concentrations evaluated in our study and CRC risk contrast with our hypothesis that these AGEs contribute to colorectal carcinogenesis. Specifically, for CML it is noteworthy that Jiao et al. (22) also reported similar inverse associations as observed in our study. However, their study was based on a sub-population of male, Finnish smokers and applied an ELISA-based methodology to assess relative differences of CML between cancer cases and controls. Unfortunately, ELISA methods have limited reproducibility and do not differentiate between protein-bound and free fractions of circulating AGEs. Our initial hypothesis for positive AGEs-CRC risk association was based on evidence from experimental studies indicating that AGEs

are DNA-damaging and can directly induce sustained inflammation in colon tissues through binding with RAGE (17,19,31). We chose to analyze protein-bound AGEs because these are specifically recognized by RAGE (32-34).

We assessed CML, CEL and MG-H1 because they are thought to be the most abundant AGEs in the body and although there is evidence that they have harmful pro-inflammatory and pro-oxidative effects, their relationship with the colonic mucosa may be very complicated. The colonic mucosa may be exposed to a variety of other AGEs, both exogenously from the diet and those produced endogenously within the body and colonic milieu. Similar to endogenous AGEs, dietary AGEs may increase the body AGEs pool, interact directly with the colon mucosal and increase the risk for obesity (35). Some researchers have classified AGEs into two categories of “non-toxic” and “toxic” (36-38). This categorization still requires much further evidence, but it is noteworthy that we have previously shown an increased risk of rectal cancer with higher circulating levels of glyceraldehyde-derived AGEs (39), that have been categorized as “toxic” (6,36). Another speculative explanation of our observations is that circulating AGEs concentration may not be reflective of their levels in colorectal tissues where they may accumulate. There is a paucity of supportive data for this assertion, and little is known about whether the concentration and actions of specific AGEs may differ between tissues. Van Heijst et al. (40) observed varying AGEs levels from an immunohistochemical expression of CML and the MGO-derived AGE argpyrimidine in various human tumours (muscle, colon, breast and larynx) suggesting that separate AGEs impact tissues differently. Therefore, relevant studies focusing on the functions, effects, and the interactions of CML, CEL and MG-H1 and additional AGEs within normal and cancerous colonic tissues are warranted.

The positive association observed with the ratio of MGO-derived AGEs:GO-derived AGEs and CRC risk suggests that the conditions that may lead to more MGO-AGEs vs. GO-AGEs may be important in CRC development. This result may implicate that higher circulating MGO versus GO may be of greater importance in CRC development than AGEs. Both MGO and GO are mainly detoxified through the glyoxalase (GLO) system and by other enzymes such as aldo-keto reductases and dicarbonyl and L-xylulose reductase (41,42). Compared to GO, MGO is more reactive (43) but is rapidly and efficiently detoxified, mainly in the liver (42). This may possibly explain the higher CRC

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

accurately and precisely measure specific AGEs in both free and protein-bound forms. The major known drawback of using UPLC-MS/MS is its relatively higher cost and the necessity for trained personnel (23,50). Additional strengths include the prospective design, the large sample size, and the ability to conduct analyses stratified by sex, and by anatomical subsite (colon vs rectum). A limitation to our study is that we lack information on other AGEs produced from MGO including other MGO-derived hydroimidazolone (MG-Hs) such as MG-H2, MG-H3 and MG-H4, Nε-(4-carboxy-4,6-dimethyl-5,6-dihydroxy-1,4,5,6-tetrahydropyrimidin-2-yl)-ornithine (THP), argpyrimidine and crosslinking dimer MOLD (8), which may all have roles in CRC development. Another limitation is the use of plasma AGEs levels which are dependent on kidney and liver functions; hence they may not represent tissue levels. Further research is required to determine how circulating AGEs measures in the same individual may relate to levels in colon tumour and normal colon tissues. Therefore, our assessment of AGEs in CRC development is far from complete, even though we analysed three major AGEs compounds. Also, our findings show that circulating measures of AGEs are likely to have differential associations with CRC, indicating that their posited detrimental properties may not be equivalent or that they vary in their pro-inflammatory capacity. More study is required on the individual and interactive roles of AGEs in the development of cancers and other chronic diseases. A deeper assessment of the qualitative pathways of AGEs production and their cumulative roles in cancer development may shed more insight into this fascinating topic. An additional limitation of this nested case-control study is the fact that blood samples and lifestyle factors were collected at baseline and may not necessarily reflect changes over time.

In conclusion, in this large, comprehensive prospective study CML and MG-H1 are inversely associated with CRC risk, contrary to our initial hypothesis. However, we observed a significantly higher CRC risk with higher ratio of MGO-derived:GO-derived AGEs. Our observations highlight the complexity of the proposed roles of AGEs in CRC development and suggest that AGEs levels may not be interpreted alone, but in consideration of their chemical origins. Additional studies examining toxic dicarbonyl AGEs precursor compounds in CRC development, and assessing the role of AGEs in the colonic milieu and within normal and tumorigenic colonic tissues are required. In addition, the 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

Grant support: Funding [WCRF 2015/1391, PI: M. Jenab] was obtained from Wereld Kanker Onderzoek Fonds (WKOF), as part of the World Cancer Research Fund International grant programme.

Acknowledgement: The authors would like to thank the EPIC study participants and staff for their valuable contribution to this research. The authors would also like to thank Mr Bertrand Hemon for the preparation of the databases. The coordination of EPIC is financially supported by International Agency for Research on Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London which has additional infrastructure support provided by the NIHR Imperial Biomedical Research Centre (BRC). The national cohorts are supported by: Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Éducation Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), Federal Ministry of Education and Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy, Compagnia di SanPaolo and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS) - Instituto de Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, and the Catalan Institute of Oncology - ICO (Spain); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford). (United Kingdom), the National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands. The EPIC-Norfolk study (DOI 10.22025/2019.10.105.00004) has received funding from the Medical Research Council (MR/N003284/1, MC-PC_13048 and MC-UU_12015/1). We are grateful to all the participants who have been part of the project and to the many members of the study teams at the University of Cambridge who have enabled this research. We are grateful to and acknowledge participants and collaborators from EPIC centres in France, Utrecht, Asturias, and Umea who also contributed data in this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest Statement: The authors indicate no conflicts of interest.

396 **References**

- 397 1. Bray, F., *et al.* (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and
398 mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.*
- 399 2. Safiri, S., *et al.* (2019) The global, regional, and national burden of colorectal cancer and its
400 attributable risk factors in 195 countries and territories, 1990–2017: a systematic analysis for
401 the Global Burden of Disease Study 2017. *The Lancet Gastroenterology & Hepatology*, **4**,
402 913-933.
- 403 3. Murphy, N., *et al.* (2019) Lifestyle and dietary environmental factors in colorectal cancer
404 susceptibility. *Mol Aspects Med.*
- 405 4. Mehta, R.S., *et al.* (2017) Dietary Patterns and Risk of Colorectal Cancer: Analysis by Tumor
406 Location and Molecular Subtypes. *Gastroenterology*, **152**, 1944-1953.e1.
- 407 5. Cerami, C., *et al.* (1997) Tobacco smoke is a source of toxic reactive glycation products. *Proc*
408 *Natl Acad Sci U S A*, **94**, 13915-20.
- 409 6. Takeuchi, M., *et al.* (2000) Immunological evidence that non-carboxymethyllysine advanced
410 glycation end-products are produced from short chain sugars and dicarbonyl compounds in
411 vivo. *Mol Med*, **6**, 114-25.
- 412 7. Scheijen, J., *et al.* (2016) Analysis of advanced glycation endproducts in selected food items
413 by ultra-performance liquid chromatography tandem mass spectrometry: Presentation of a
414 dietary AGE database. *Food Chem*, **190**, 1145-1150.
- 415 8. Sharma, C., *et al.* (2015) Advanced glycation End-products (AGEs): an emerging concern for
416 processed food industries. *J Food Sci Technol*, **52**, 7561-76.
- 417 9. Schalkwijk, C.G., *et al.* (2020) Methylglyoxal, a Highly Reactive Dicarbonyl Compound, in
418 Diabetes, Its Vascular Complications, and Other Age-Related Diseases. *Physiological Reviews*,
419 **100**, 407-461.
- 420 10. Turk, Z. (2010) Glycotoxines, carbonyl stress and relevance to diabetes and its complications.
421 *Physiol Res*, **59**, 147-56.
- 422 11. Thornalley, P.J. (2005) Dicarbonyl intermediates in the maillard reaction. *Ann N Y Acad Sci*,
423 **1043**, 111-7.
- 424 12. Nowotny, K., *et al.* (2018) Dietary advanced glycation end products and their relevance for
425 human health. *Ageing Res Rev*, **47**, 55-66.
- 426 13. Arribas-Lorenzo, G., *et al.* (2010) Analysis, Distribution, and Dietary Exposure of Glyoxal and
427 Methylglyoxal in Cookies and Their Relationship with Other Heat-Induced Contaminants.
428 *Journal of Agricultural and Food Chemistry*, **58**, 2966-2972.
- 429 14. Raupbach, J., *et al.* (2020) Proteasomal degradation of glycated proteins depends on
430 substrate unfolding: Preferred degradation of moderately modified myoglobin. *Free Radical*
431 *Biology and Medicine*, **152**, 516-524.
- 432 15. Palanissami, G., *et al.* (2018) RAGE and Its Ligands: Molecular Interplay Between Glycation,
433 Inflammation, and Hallmarks of Cancer-a Review. *Horm Cancer*, **9**, 295-325.
- 434 16. Sakellariou, S., *et al.* (2016) Clinical significance of AGE-RAGE axis in colorectal cancer:
435 associations with glyoxalase-I, adiponectin receptor expression and prognosis. *BMC Cancer*,
436 **16**, 174.
- 437 17. Kuniyasu, H., *et al.* (2003) Differential effects between amphoterin and advanced glycation
438 end products on colon cancer cells. *Int J Cancer*, **104**, 722-7.
- 439 18. Chen, H., *et al.* (2014) Advanced glycation end products increase carbohydrate responsive
440 element binding protein expression and promote cancer cell proliferation. *Mol Cell*
441 *Endocrinol*, **395**, 69-78.
- 442 19. Shimomoto, T., *et al.* (2012) Advanced glycation end products (AGE) induce the receptor for
443 AGE in the colonic mucosa of azoxymethane-injected Fischer 344 rats fed with a high-linoleic
444 acid and high-glucose diet. *J Gastroenterol*, **47**, 1073-83.
- 445 20. Wang, P., *et al.* (2018) Advanced Glycation End Products Increase MDM2 Expression via
446 Transcription Factor KLF5. *J Diabetes Res*, **2018**, 3274084.

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

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

FIGURES LEGENDS

Figure 1: Schematic representation of the formation of the AGEs and the rationale for the calculation of the ratios

Abbreviations: AGE, advanced glycation end-product; CML, N^ε-carboxy-methyllysine; CEL, N^ε-carboxy-ethyllysine; GO, glyoxal; MG-H1, N^δ-(5-hydro-5-methyl-4-imidazolone-2-yl)-ornithine; MGO, methylglyoxal

AGEs are absorbed from the diet or formed during the Maillard reactions from the Amadori or Heyns products and from the glycating actions of dicarbonyls such as MGO and GO. CML is derived from GO whereas MG-H1 and CEL are derived from MGO.

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

	Cases (1,378)	Controls (n=1,378)
Women, %	51.7	51.5
Anthropometry, mean (SD)		
BMI, kg/m ²	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	<i>P</i> _{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	

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-imidazol-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		$P_{\text{heterogeneity}}$ 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
$P_{\text{heterogeneity}}$ 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
$P_{\text{heterogeneity}}$ 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
$P_{\text{heterogeneity}}$ by sex		0.313		0.332		0.701	
Σ AGEs, 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
$P_{\text{heterogeneity}}$ 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
$P_{\text{heterogeneity}}$ 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
<i>P</i> _{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-imidazol-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)	<i>P</i> for heterogeneity	Yes (n=247)	No (n=1131)	<i>P</i> 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-imidazolone-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)

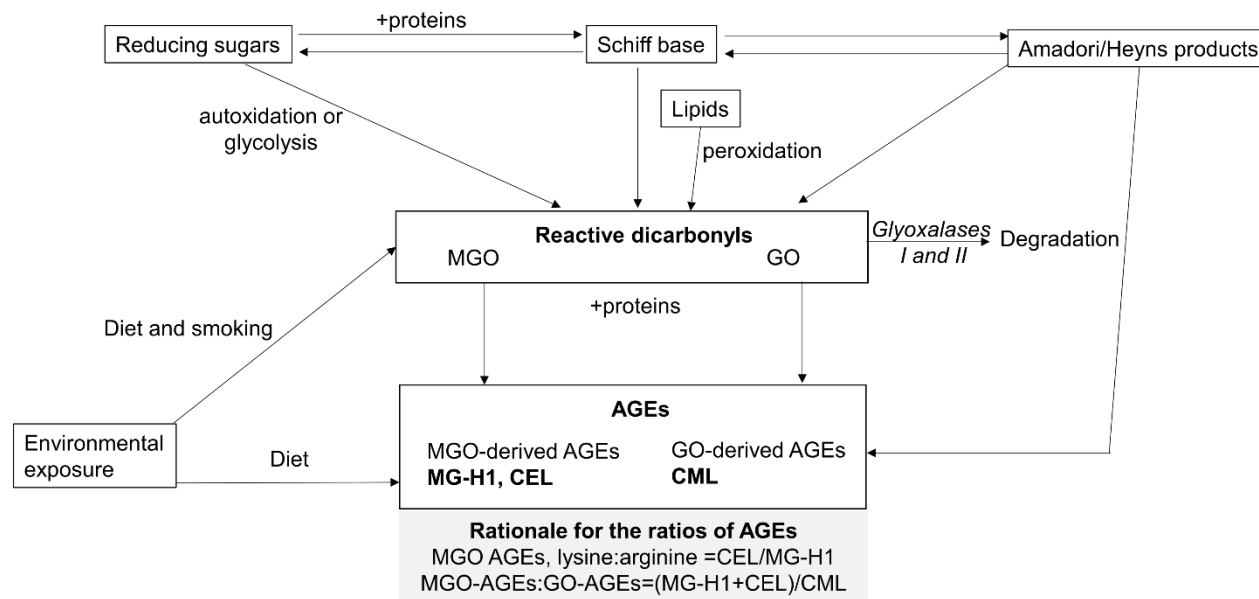


Figure 1