



OPEN

## Biomarker-estimated flavan-3-ol intake is associated with lower blood pressure in cross-sectional analysis in EPIC Norfolk

Javier I. Ottaviani<sup>1</sup>, Abigail Britten<sup>2</sup>, Debora Lucarelli<sup>2</sup>, Robert Luben<sup>3</sup>, Angela A. Mulligan<sup>2</sup>, Marleen A. Lentjes<sup>4</sup>, Reedmond Fong<sup>5</sup>, Nicola Gray<sup>6</sup>, Philip B. Grace<sup>7</sup>, Deborah H. Mawson<sup>7</sup>, Amy Tym<sup>7</sup>, Antonia Wierzbicki<sup>7</sup>, Nita G. Forouhi<sup>2</sup>, Kay-Tee Khaw<sup>3</sup>, Hagen Schroeter<sup>1</sup> & Gunter G. C. Kuhnle<sup>6</sup>✉

Flavan-3-ols are a group of bioactive compounds that have been shown to improve vascular function in intervention studies. They are therefore of great interest for the development of dietary recommendation for the prevention of cardio-vascular diseases. However, there are currently no reliable data from observational studies, as the high variability in the flavan-3-ol content of food makes it difficult to estimate actual intake without nutritional biomarkers. In this study, we investigated cross-sectional associations between biomarker-estimated flavan-3-ol intake and blood pressure and other CVD risk markers, as well as longitudinal associations with CVD risk in 25,618 participants of the European Prospective Investigation into Cancer (EPIC) Norfolk cohort. High flavan-3-ol intake, achievable as part of an habitual diet, was associated with a significantly lower systolic blood pressure (− 1.9 (− 2.7; − 1.1) mmHg in men and − 2.5 (− 3.3; − 1.8) mmHg in women; lowest vs highest decile of biomarker), comparable to adherence to a Mediterranean Diet or moderate salt reduction. Subgroup analyses showed that hypertensive participants had stronger inverse association between flavan-3-ol biomarker and systolic blood pressure when compared to normotensive participants. Flavanol intake could therefore have a role in the maintenance of cardiovascular health on a population scale.

High blood pressure is a leading disease risk factor globally, and cardiovascular disease (CVD) is a main cause of death<sup>1</sup>. Therefore, the primary prevention of CVD remains of utmost importance, and changes in dietary factors have an important rôle. While dietary recommendations for primary disease prevention have focused mainly on dietary patterns and macronutrients<sup>2</sup>, increasing attention has been given to a group of non-nutritive dietary compounds, bioactives, that are thought to exert a physiological effect and to modulate disease risk<sup>3</sup>. Flavan-3-ols are a major class of dietary bioactives<sup>4–6</sup>, belonging to the group of polyphenolics, commonly found in tea, pome fruits, berries, cocoa-derived products and nuts. Accumulating evidence from dietary intervention studies shows that the intake of flavan-3-ols improves vascular function in healthy adults<sup>6,7</sup>. Indeed, multiple clinical dietary intervention studies have demonstrated flavanol-intake related cardiovascular health benefits by assessing physiological endpoints including blood pressure, flow-mediated arterial dilation, augmentation index, pulse wave velocity and arterial stiffness, as well as atherogenesis<sup>8–10</sup>. However, the data currently available are neither describing effects at sufficient scale nor were derived from long-term investigations and are thus insufficient provide the basis for population-based dietary guidance<sup>6</sup>. While ongoing large-scale clinical dietary intervention studies, such as COSMOS (NCT02422745<sup>11</sup>), with specific focus on cardiovascular disease risk and outcomes measures that include stroke, myocardial infarction and blood pressure, are intended to close this gap, large-scale observational studies can provide crucial information about associations between habitual flavan-3-ol

<sup>1</sup>Mars, Inc., McLean, VA, USA. <sup>2</sup>MRC Epidemiology Unit, University of Cambridge, Cambridge, UK. <sup>3</sup>Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. <sup>4</sup>School of Medical Sciences, Örebro University, Örebro, Sweden. <sup>5</sup>Department of Nutrition, UC Davis, Davis, CA, USA. <sup>6</sup>Department of Food and Nutritional Sciences, University of Reading, Reading, UK. <sup>7</sup>LGC, Newmarket Road, Fordham, UK. ✉email: g.g.kuhnle@reading.ac.uk

intake and vascular health, especially as these studies usually include a wide range of different foods and beverages and are based on a more heterogeneous population representative of the general public.

Importantly, large-scale observational studies rely on the accurate assessment of intake. To date, all such studies aiming to estimate flavan-3-ol intake were based on self-reported dietary data, food frequency questionnaires (FFQs) or food diaries, in combination with food composition data. While this approach, in the absence of practicable alternatives, has been tacitly accepted as de facto standard<sup>12–16</sup>, it introduces significant limitations that substantially affect outcome and interpretation. Self-reported dietary data are subject to a number of limitations and have been demonstrated to be subject to systematic bias<sup>17</sup>. Detailed analyses, for example of protein and energy<sup>18</sup>, but also sugar intake<sup>19</sup>, have shown systematic under-reporting. While these methods can provide reliable data on dietary patterns and intake of individual foods, these limitations affect the ability to estimate intake of individual compounds. This is further exacerbated by the reliance on food composition data, which can only provide data on average food content and not the composition of the foods actually consumed. For flavan-3-ols but also other compounds found in foods, this approach introduces significant error due to the large variability of food composition<sup>20</sup>, the effects of processing<sup>21</sup> and differences in bioavailability. Thus, the reliance on food composition data to estimate flavan-3-ol intake introduces a considerable measurement error: for example, the amount found in tea, one of the main dietary sources in the UK diet<sup>22</sup>, ranges from 10 to 330 mg/100 g<sup>20</sup>. This problem is made worse by the common reliance on FFQs, which do not provide sufficient detail on food intake to allow an accurate estimate of actual intake<sup>23</sup>.

In contrast, nutritional biomarkers, which are assessed by measuring the systemic presence of dietary compounds or their metabolites, have the potential to mitigate the above limitations and thus enable objective and accurate estimates of actual intake<sup>24–26</sup>. As biomarkers for estimating flavan-3-ol intake were not available previously, we developed and evaluated at scale nutritional biomarkers to estimate the intake of flavan-3-ols in general, based on the flavan-3-ol-derived microbial metabolite 5-3',4'-dihydroxyphenyl- $\gamma$ -valerolactone (gVLM)<sup>27</sup>, and one specific for (–)-epicatechin intake, based on structurally related (–)-epicatechin metabolites (SREM)<sup>28</sup>. The biomarkers derived from those metabolites, referred to as gVLM<sub>B</sub> and SREM<sub>B</sub>, are surrogate biomarkers<sup>29</sup> when assessed in spot urine and therefore allow people to be ranked according to their flavan-3-ol intake.

These novel biomarkers allow rigorous, and more objective and accurate investigations into associations between actual flavan-3-ol intake and health in observational cohorts at scales relevant to human populations. The primary objective of this study was to investigate cross-sectional associations between biomarker-estimated flavan-3-ol intake and blood pressure in more than 25,000 participants of the Norfolk cohort of the European Prospective Investigation into Cancer Study (EPIC-Norfolk). The secondary objectives of the study were to investigate cross-sectional associations with other cardio-vascular disease risk factors and prospective associations with cardio-vascular disease risk.

## Results

**Study population and biomarker.** This study was based on data of 25,618 participants (14,026 women, 55%) of EPIC-Norfolk, after the exclusion of those lost to follow-up (2 women, 1 man), and those who withdrew consent (18). Table 1 shows a summary of the baseline characteristics of the study population; more detailed data, including information on missing data, are shown as Supplemental Information (Supplementary Tables 1 and 2). Biomarker concentrations were available for 24,152 participants (13,273 women, 55%). Using specific gravity, spot urine samples were adjusted for dilution (25), and these data were available for 21,812 participants (11,974 women, 55%), as specific gravity data were not available for all. Missing data were assumed to be missing at random and imputed using multiple imputations. The concentrations of both biomarkers, gVLM<sub>B</sub> for flavan-3-ols and SREM<sub>B</sub> for (–)-epicatechin, were correlated (Pearson's  $\rho = 0.45$ ).

We have investigated the correlation between biomarker-estimated flavan-3-ol and self-reported food and flavan-3-ol intake. There were weak correlations ( $R^2 < 0.2$ ) between flavan-3-ol biomarkers and the consumption of foods associated with flavan-3-ol intake in the diet of participants in EPIC-Norfolk (Fig. 1), and virtually no correlation with flavan-3-ol intake estimated using 7-day food diaries ( $R^2$ : 0.01 for gVLM<sub>B</sub> and 0.07 for SREM<sub>B</sub>, adjusted for energy intake). These findings are consistent with the limitations of estimating bioactive-intake from self-reported dietary data described above, in particular the high variability in food composition.

**Cross-sectional associations between flavan-3-ol biomarker and blood pressure.** The evidence available from a wide range of clinical intervention studies show that flavan-3-ols can have a vasculoprotective effect<sup>6</sup>, but this has never been shown on a large scale in a general population. We have therefore investigated the cross-sectional association between flavan-3-ol intake and blood pressure at baseline (Fig. 2). Table 2 shows consistent inverse associations between biomarkers of flavan-3-ol (gVLM<sub>B</sub>) and (–)-epicatechin (SREM<sub>B</sub>) intake and systolic blood pressure in all models tested. We have further investigated whether the biomarker merely acts as a surrogate marker of specific dietary patterns that are associated with blood pressure. Tea is a main dietary source of flavan-3-ol in EPIC-Norfolk<sup>15</sup>, and thus gVLM<sub>B</sub> and SREM<sub>B</sub> could both act as a marker of high tea or low coffee intake, as there was a strong inverse association between tea and coffee intake (Pearson's  $\rho = 0.4$ ); similarly, fruits and vegetables can be an important contributor of flavan-3-ols. However, adjusting our data analysis additionally for tea and coffee intake, as well plasma vitamin C, as a surrogate marker of fruit and vegetable intake<sup>30</sup>, associations did not change materially (Table 2). When using gVLM<sub>B</sub> as biomarker of the intake of flavan-3-ol in general, the difference in systolic blood pressure between bottom and top decile of biomarker concentrations, the median of the bottom and top quintile, was  $-1.9$  ( $-2.7$ ;  $-1.1$ ) mmHg in men and  $-2.5$  ( $-3.3$ ;  $-1.8$ ) mmHg in women (model 5). Compared to the results using gVLM<sub>B</sub>, the differences in blood pressure between the bottom and top decile were larger when using SREM<sub>B</sub> as specific biomarker of (–)-epicatechin intake,  $-2.4$  ( $-3.3$ ;  $-1.5$ ) mmHg in men and  $-2.5$  ( $-3.6$ ;  $-2.0$ ) mmHg in women. We found similar associa-

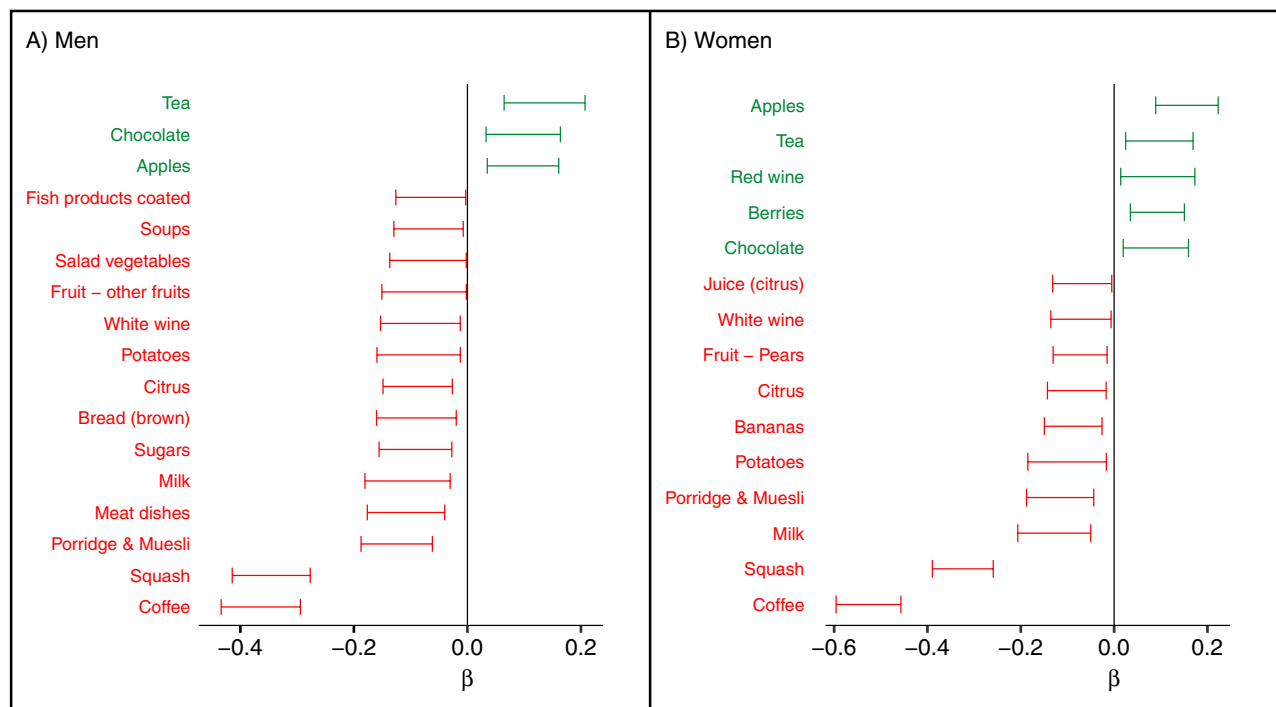
	All	Men	Women
n	25,618	11,592	14,026
Age (years)	58.7 (9.3)	59.1 (9.3)	58.4 (9.3)
BMI (kg/m <sup>2</sup> )	26.4 (3.9)	26.5 (3.3)	26.2 (4.4)
Female (%)	14,026 (55%)	–	–
<b>Physical activity</b>			
Inactive	7853 (31%)	3579 (31%)	4274 (31%)
Moderately inactive	7344 (29%)	2853 (25%)	4491 (32%)
Moderately active	5773 (23%)	2657 (23%)	3116 (22%)
Active	4647 (18%)	2502 (22%)	2145 (15%)
<b>Smoking status</b>			
Current	2979 (12%)	1402 (12%)	1577 (11%)
Former	10,751 (42%)	6276 (55%)	4475 (32%)
Never	11,668 (46%)	3833 (33%)	7835 (56%)
Systolic BP (mmHg)	135 (18)	137 (18)	134 (19)
Diastolic BP (mmHg)	83 (11)	84 (11)	81 (11)
Cholesterol (mmol/L)	6.2 (1.2)	6.0 (1.1)	6.3 (1.2)
LDL (mmol/L)	4.0 (1.0)	3.9 (1.0)	4.0 (1.1)
HDL (mmol/L)	1.4 (0.4)	1.2 (0.3)	1.6 (0.4)
Triglycerides (mmol/L)	1.8 (1.1)	2.1 (1.2)	1.6 (1.0)
cRP (mg/L)	3.1 (6.3)	3.0 (6.0)	3.0 (5.9)
<b>Medication user</b>			
Lipid lowering drugs	377 (2%)	173 (2%)	204 (2%)
Anti-hypertensive drugs	4798 (19%)	2165 (19%)	2633 (19%)
<b>Biomarker</b>			
gVLM (μmol/L)	9.5 (16.0)	10.5 (16.5)	8.7 (15.5)
gVLM (μmol/L) <sup>a</sup>	9.3 (15.7)	10.2 (16.2)	8.6 (15.3)
SREM <sub>B</sub> (μmol/L)	2.0 (3.0)	2.2 (3.0)	1.8 (3.0)
SREM <sub>B</sub> (μmol/L) <sup>a</sup>	1.9 (2.9)	2.1 (2.9)	1.8 (2.9)
<b>Disease incidence<sup>b</sup></b>			
All CVD	13,969 (55%)	6907 (60%)	7062 (50%)
<b>Mortality<sup>b</sup></b>			
All cause	8030 (31%)	4277 (37%)	3753 (27%)
CVD	2613 (10%)	1474 (13%)	1139 (8%)

**Table 1.** Baseline characteristics and disease incidence of 25,618 participants of EPIC Norfolk. Data shown are mean (SD) or absolute number and proportion. More details, including number of missing data, are shown in Supplementary Tables 1 and 2. <sup>a</sup>Adjusted by specific gravity. <sup>b</sup>Number at end of follow-up.

tions for diastolic blood pressure and biomarker-estimated flavan-3-ol intake, with a difference between the bottom and top decile of approximately 1 mmHg.

**Association between flavanol intake and other cardiovascular disease risk markers.** We investigated cross-sectional associations between flavan-3-ol and (–)-epicatechin biomarker and other established CVD risk markers (blood lipids and c-reactive protein). There were small differences in blood lipids between the bottom and top decile of flavan-3-ol intake assessed with both SREM<sub>B</sub> and gVLM<sub>B</sub>, with participants in the top decile of biomarker having lower blood cholesterol and LDL concentrations, but higher triglycerides. The associations were very similar for SREM<sub>B</sub> and gVLM<sub>B</sub>. We did not find any associations of the flavanol biomarkers with c-reactive protein concentration (Table 3).

**Association with CVD incidence and mortality.** High flavan-3-ol intake was associated with lower blood pressure and an overall better blood lipid profile and may therefore tenably affect overall CVD risk and mortality. Thus, in a secondary analysis, we have investigated associations between gVLM<sub>B</sub> as biomarker of flavan-3-ol intake and CVD risk and CVD and all-cause mortality. After a median of 19.5 (IQR 17.9–20.9) years of follow-up, 8030 (31%) participants had died and 13,969 (55%) had developed a cardiovascular disease. Overall, there were no consistent, statistically significant associations between flavan-3-ol biomarker and CV disease incidence or all cause or CVD mortality (Table 3).



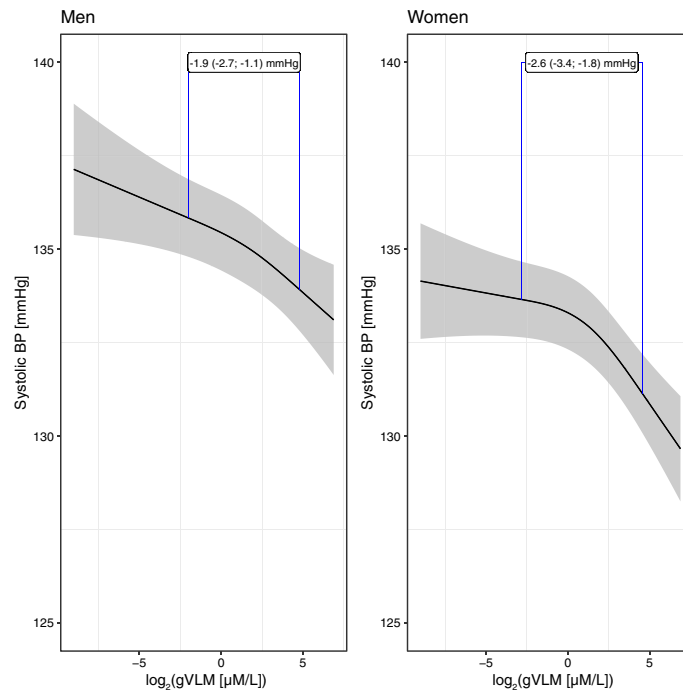
**Figure 1.** Association between biomarker-estimated flavan-3-ol intake ( $gVLM_B$ ) and selected food groups from 7-day diaries. Biomarker-estimated intake was positively associated with tea, wine and apple intake, whereas there were inverse associations with coffee and squash (cordial) intake.  $\beta$  are changes in specific gravity-adjusted biomarker concentration per SD change of respective reported food group weight, adjusted by total energy intake, social class, BMI and age; only food groups with statistically significant association ( $p < 0.05$ ) are shown.

**Subgroup and sensitivity analyses.** We conducted sub-group and sensitivity analyses to further investigate the associations between biomarker-estimated flavan-3-ol intake and blood pressure (Fig. 3). We aimed at investigating whether or not pre-existing CVD or CVD risk would affect the association between systolic blood pressure and biomarker-estimated flavan-3-ol intake. Therefore, we compared the difference in systolic blood pressure between the top and bottom decile of the flavan-3-ol biomarker in participants with and without pre-existing CVD or CVD risk (age, hypertension, overweight, and prevalent CVD at baseline). Hypertensive participants had stronger inverse association between flavan-3-ol biomarker and systolic blood pressure when compared to normotensive participants, in particular when using  $SREM_B$  as biomarker. Similar differences were observed between hypertensive and normotensive women using  $gVLM_B$ -estimated flavan-3-ol intakes. In addition, the difference in systolic blood pressure between the top and bottom decile of the  $SREM_B$ - and  $gVLM_B$ -estimated flavan-3-ol intake in older men (>60 years) was greater than that in younger men. Finally, significant differences were also observed in women with low and high risk of CVD using  $SREM_B$  as biomarker of flavan-3-ol intake.

## Discussion

In participants of EPIC-Norfolk, a large cohort representative of the older general public in England<sup>31</sup>, high flavan-3-ol intake was associated with a significantly lower systolic and diastolic blood pressure and was inversely associated with blood lipids. We did not observe consistent, statistically significant associations between biomarkers of flavan-3-ol intake and CVD incidence or mortality.

This study was enabled by recently developed nutritional biomarkers,  $gVLM_B$  and  $SREM_B$ , that allow for estimating specifically the intake of flavan-3-ols. We have shown previously that  $gVLM_B$  and  $SREM_B$  reflect actual intake of flavan-3-ols and (-)-epicatechin respectively, yet importantly here we found only a weak correlation between biomarker and self-reported food intake<sup>27,28</sup>. This finding strongly supports our previous findings of the impact of the high variability in food composition on self-reported dietary assessment methods<sup>23</sup>. The high variability in food composition of many of the main sources of dietary flavan-3-ol make an accurate estimate without an analysis of the actual food consumed virtually impossible. For example, food composition data<sup>20</sup> for black tea give a range of flavan-3-ol content of 3–64 mg/100 mL, and thus five cups of tea can contain between 23 and 480 mg of flavan-3-ols. Thus, a person consuming a single cup of tea with high flavan-3-ol content consumes considerably more flavan-3-ols than a person consuming five cups of tea with low flavan-3-ol content. A wide variability in food composition has been reported for other foods, too. Indeed, even for foods grown on the self-same plant, up to 2.5-fold differences for some nutrients have been observed<sup>32</sup>. In contrast, the objective nutritional biomarkers used in this study take into account the diversity in the foods consumed and differences in bioavailability as they rely on the systemic presence of the respective compounds. The inclusion of the main dietary sources of flavan-3-ols and (-)-epicatechin in our statistical models (i.e. tea as well as fruit



**Figure 2.** Association between biomarker estimated flavan-3-ol intake and systolic blood pressure. Predicted systolic blood pressure (95% confidence interval) in men (left,  $n=11,592$ ) and women (right,  $n=14,026$ ) adjusted for (model 5) age and BMI, smoking status, physical activity and social class, plasma vitamin C as marker of fruit and vegetable intake, tea and coffee intake, baseline health (self-reported diabetes mellitus, myocardial infarction, cerebrovascular accident), family history of myocardial infarction, use of anti-hypertensive or lipid-lowering drugs and menopausal status and hormone replacement therapy in women. The blue lines indicate the estimated differences in systolic blood pressure between low (10th percentile) and high (90th percentile) concentrations of the biomarker. Biomarker concentrations were adjusted by specific gravity.

and vegetables<sup>22</sup>) did not affect outcomes (Table 2, models 3 and 5), which strongly supports the notion that the data provided by the nutritional biomarkers is related to the actual intake of bioactive compound and does not just reflect dietary patterns.

The two different biomarkers of flavan-3-ol intake used here ( $gVLM_B$  and  $SREM_B$ ), enabled estimating the dietary intake of flavan-3-ols<sup>27</sup> and the flavan-3-ol (–)-epicatechin<sup>28</sup>, respectively. The differences between the effect sizes estimated with either biomarker were negligible (Tables 2 and 3), and this can to some extent be explained by the correlation between the two biomarkers. However, the two biomarkers used in this study originate from two distinct metabolic pathways,  $gVLM_B$  from the catabolism of flavan-3-ols in the gut microbiome,  $SREM_B$  from phase II biotransformation reactions of (–)-epicatechin. Thus, these results pose the question as to whether the observed association between flavan-3-ol intake and blood pressure could be explained by the specific intake of (–)-epicatechin, as it is the only compound measured with both biomarkers. In this context, previous intervention studies have established a vascular effect for (–)-epicatechin<sup>33</sup>, and showed that only (–)-epicatechin, but no other type of flavan-3-ols exert such activity<sup>34</sup>. While larger controlled dietary intervention studies are necessary to establish the actual bioactive compound and mode of action, our results clearly show an inverse association between flavan-3-ols, including (–)-epicatechin, and blood pressure, and thereby contribute to the data available to investigate a causal effect. While the specific molecular mechanisms that underlie the cardiovascular effects of flavanols are still under investigation, currently published work in this context indicates that flavanols mediate a range of effects on the cellular-/molecular level that impact on endothelial function, nitric oxide-dependent arterial dilation, thrombogenic responses, and processes related to vascular inflammation, angiogenesis, and endothelial repair<sup>8–10,35–37</sup>.

In comparison with the observed associations with blood pressure, the associations with blood lipids were rather modest. However, they were of a similar magnitude and direction than those observed in the Minnesota Green Tea Trial (MGTT<sup>38</sup>), even though flavan-3-ol intakes in EPIC-Norfolk were considerably lower.

**Impact on health.** In this study, we could demonstrate significant inverse associations between biomarkers of flavan-3-ol and (–)-epicatechin intake and blood pressure at baseline. However, there were no consistent, statistically significant associations with CVD risk or mortality (CVD related and all-cause) (Table 4). This can be explained by the magnitude of difference in systolic blood pressure observed, which would not be expected to have significant impact on individual CVD risk (approximately 0.2 percentage points reduction in 10-year CVD risk based on QRISK 3<sup>39</sup>).

	Men (n = 11,592)		Women <sup>b</sup> (n = 14,026)	
	Systolic BP (mm Hg)	Diastolic BP (mm Hg)	Systolic BP (mm Hg)	Diastolic BP (mm Hg)
<b>gVLM<sub>B</sub><sup>c</sup></b>				
Model 0 <sup>a</sup>	-1.6 (-2.4; -0.8)	-1.0 (-1.5; -0.4)	-1.8 (-2.5; -1.0)	-1.1 (-1.6; -0.6)
Model 1	-1.8 (-2.6; -1.0)	-1.2 (-1.7; -0.7)	-2.2 (-3.0; -1.5)	-1.4 (-1.9; -0.9)
Model 2	-1.8 (-2.7; -1.0)	-1.2 (-1.7; -0.6)	-2.3 (-3.0; -1.5)	-1.4 (-1.9; -0.9)
Model 3	-2.0 (-2.8; -1.2)	-1.3 (-1.8; -0.7)	-2.6 (-3.3; -1.8)	-1.6 (-2.0; -1.1)
Model 4	-1.8 (-2.6; -1.0)	-1.2 (-1.7; -0.7)	-2.3 (-3.0; -1.5)	-1.4 (-1.9; -0.9)
Model 5	-1.9 (-2.7; -1.1)	-1.3 (-1.8; -0.7)	-2.5 (-3.3; -1.8)	-1.6 (-2.0; -1.1)
<b>SREM<sub>B</sub><sup>c</sup></b>				
Model 0	-1.8 (-2.6; -0.9)	-0.7 (-1.3; -0.2)	-1.3 (-2.1; -0.5)	-0.3 (-0.8; 0.2)
Model 1	-2.0 (-2.8; -1.2)	-0.9 (-1.4; -0.4)	-1.9 (-2.7; -1.1)	-0.8 (-1.2; -0.3)
Model 2	-2.1 (-2.9; -1.2)	-0.9 (-1.4; -0.4)	-2.0 (-2.7; -1.2)	-0.8 (-1.3; -0.3)
Model 3	-2.5 (-3.3; -1.6)	-1.2 (-1.7; -0.6)	-2.9 (-3.7; -2.1)	-1.2 (-1.7; -0.7)
Model 4	-2.0 (-2.8; -1.2)	-0.9 (-1.4; -0.4)	-1.9 (-2.7; -1.1)	-0.8 (-1.3; -0.3)
Model 5	-2.4 (-3.3; -1.5)	-1.2 (-1.7; -0.6)	-2.8 (-3.6; -2.0)	-1.2 (-1.7; -0.6)

**Table 2.** Associations between biomarker-estimated flavan-3-ol intake and blood pressure. Results shown are estimated differences (95% CI) between low (10th percentile) and high (90th percentile) biomarker concentrations, using multi-variable linear regression different statistical models. <sup>a</sup>Model 0: adjusted for age; model 1: additionally adjusted for BMI; model 2: additionally adjusted for smoking status, physical activity and social class; model 3: additionally adjusted plasma vitamin C as marker of fruit and vegetable intake, tea and coffee intake; model 4: model 2, additionally adjusted for baseline health (self-reported diabetes mellitus, myocardial infarction, cerebrovascular accident), family history of myocardial infarction, use of anti-hypertensive or lipid-lowering drugs; model 5: model 3, additionally adjusted for baseline health (self-reported diabetes mellitus, myocardial infarction, cerebrovascular accident), family history of myocardial infarction, use of anti-hypertensive or lipid-lowering drugs. <sup>b</sup>Additionally adjusted for menopausal status and hormone replacement therapy. <sup>c</sup>Biomarker concentrations were adjusted by specific gravity.

The difference in systolic blood pressure observed here between low and high biomarker concentration in the cross-sectional analysis (approximately 2 mmHg) is similar to the reduction in blood pressure observed in dietary intervention studies<sup>7</sup>. This difference is comparable to those observed with a Mediterranean diet in the PREDIMED trial (1.5 mmHg<sup>40</sup>) or a moderate reduction in salt intake in the DASH-Sodium trial (2.1 mmHg, high to intermediate sodium intake<sup>41</sup>), and could have considerable impact on health at a population scale. However, large-scale dietary intervention studies such as COSMOS (NCT02422745<sup>11</sup>) are required to confirm whether the observed differences in blood pressure can be explained by differences in flavan-3-ol intake.

A subgroup analyses showed that the association between intake and blood pressure was strongest among participants at higher risk of developing cardiovascular diseases, in particular older participants and those with existing hypertension, confirming results from previous small-scale dietary intervention studies<sup>7</sup>. The association between intake and blood pressure therefore follows a progressive model<sup>42</sup>, where the strongest effect size is found in those with higher blood pressure (Fig. 4). Such a model has also been observed in other dietary interventions such as the effect of potassium<sup>43</sup> or sodium<sup>44</sup> intake on blood pressure or the DASH diet<sup>45</sup>. In such a model, even a small reduction in blood pressure can have a considerable impact on morbidity and mortality on a population scale as it reduces the prevalence of hypertension and pre-hypertension, and thus the number of people at higher risk of CVD. Indeed, a reduction of 3 mmHg systolic blood pressure can be translated into a reduction in all-cause mortality by 3%<sup>42,46</sup>. Flavanol intake could therefore have a role in the maintenance of cardiovascular health on a population scale.

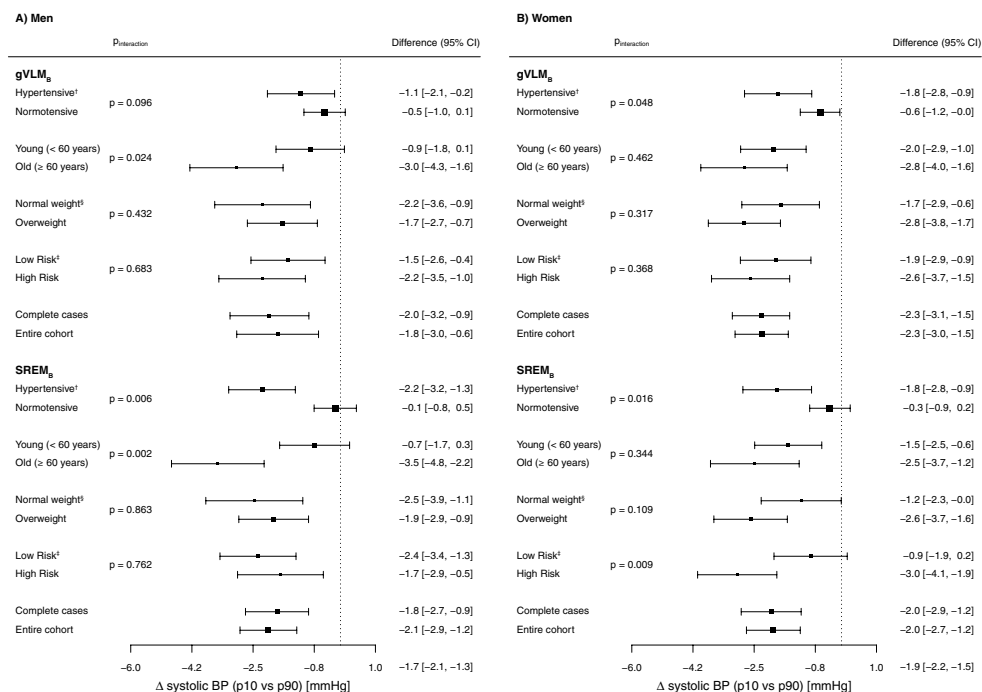
**Nutritional interpretation.** The differences in blood pressure observed in this study were seen in a population with relatively high habitual intakes of flavan-3-ol and not any supplementation as has been used in intervention studies. In this context, it is noteworthy that, while representative of the UK population, the diet of participants of EPIC-Norfolk regarding flavan-3-ol intake may not be representative of countries without a strong tea culture, including countries in continental Europe and the US<sup>15</sup>. Compared to EPIC-Norfolk, it is therefore possible that a larger fraction of the population in countries without a tea culture could benefit from increasing the intake of flavan-3-ol in their diets. In addition, the question is still open as to whether or not the intake of flavan-3-ols beyond the ranges in the EPIC-Norfolk population would be associated with a further reduction in blood pressure. In any case, the large variability of flavan-3-ol content in foods precludes the identification of a diet that would result in a given increase of flavan-3-ol intake. While incorporating foods and beverages, such as tea, apples and berries, would probably increase intake of flavan-3-ols in the diet, it will depend not only on the type of food, but also on the actual product variety (species) consumed, the manufacturing conditions in which that product was generated and preserved and mode of preparation. This makes more specific recommendations impossible.

	Cholesterol (mmol/L)	HDL (mmol/L)	LDL (mmol/L)	Triglycerides (mmol/L)	CRP (mg/L)
<b>Men (n = 11,592)</b>					
<b>gVLM<sub>B</sub><sup>c</sup></b>					
Model 0 <sup>a</sup>	-0.1 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.0; 0.1)	-0.1 (-0.5; 0.2)
Model 1	-0.1 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.0; 0.1)	-0.2 (-0.5; 0.2)
Model 2	-0.1 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.0; 0.1)	-0.2 (-0.5; 0.1)
Model 3	-0.1 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.0 (0.0; 0.1)	-0.2 (-0.6; 0.1)
Model 4	-0.1 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.0; 0.1)	-0.2 (-0.5; 0.2)
Model 5	-0.1 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.0; 0.1)	-0.2 (-0.5; 0.1)
<b>SREM<sub>B</sub><sup>c</sup></b>					
Model 0	-0.1 (-0.2; -0.1)	0.0 (0.0; 0.0)	-0.2 (-0.2; -0.1)	0.1 (0.1; 0.2)	0.2 (-0.2; 0.5)
Model 1	-0.1 (-0.2; -0.1)	0.0 (0.0; 0.0)	-0.2 (-0.2; -0.1)	0.1 (0.1; 0.2)	0.1 (-0.2; 0.5)
Model 2	-0.1 (-0.2; -0.1)	0.0 (0.0; 0.0)	-0.2 (-0.2; -0.1)	0.1 (0.1; 0.2)	0.1 (-0.3; 0.4)
Model 3	-0.1 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.2; -0.1)	0.1 (0.0; 0.2)	0.0 (-0.4; 0.3)
Model 4	-0.1 (-0.2; -0.1)	0.0 (0.0; 0.0)	-0.2 (-0.2; -0.1)	0.1 (0.1; 0.2)	0.1 (-0.3; 0.4)
Model 5	-0.1 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.2; -0.1)	0.1 (0.1; 0.2)	0.0 (-0.3; 0.4)
<b>Women<sup>b</sup> (n = 14,026)</b>					
<b>gVLM<sub>B</sub><sup>c</sup></b>					
Model 0	-0.1 (-0.1; 0.0)	0.0 (-0.1; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.1; 0.2)	0.3 (0.0; 0.6)
Model 1	-0.1 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.0; 0.1)	0.2 (-0.1; 0.5)
Model 2	-0.1 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.0; 0.1)	0.2 (-0.1; 0.5)
Model 3	-0.1 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.0; 0.1)	0.1 (-0.2; 0.4)
Model 4	-0.1 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.0; 0.1)	0.2 (-0.1; 0.5)
Model 5	-0.1 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.0; 0.1)	0.1 (-0.2; 0.4)
<b>SREM<sub>B</sub><sup>c</sup></b>					
Model 0	0.0 (-0.1; 0.1)	0.0 (0.0; 0.0)	0.0 (-0.1; 0.0)	0.2 (0.1; 0.2)	0.4 (0.1; 0.7)
Model 1	0.0 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.1; 0.2)	0.3 (0.0; 0.6)
Model 2	0.0 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.1; 0.2)	0.3 (0.0; 0.6)
Model 3	0.0 (-0.1; 0.1)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.1; 0.1)	0.1 (-0.3; 0.4)
Model 4	0.0 (-0.1; 0.0)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.1; 0.2)	0.3 (0.0; 0.6)
Model 5	0.0 (-0.1; 0.1)	0.0 (0.0; 0.0)	-0.1 (-0.1; 0.0)	0.1 (0.1; 0.1)	0.1 (-0.2; 0.4)

**Table 3.** Associations between biomarker-estimated flavan-3-ol intake and CVD risk markers. Results shown are estimated differences (95% CI) between low (10th percentile) and high (90th percentile) biomarker concentrations, using multi-variable linear regression different statistical models. <sup>a</sup>Model 0: adjusted for age; model 1: additionally adjusted for BMI; model 2: additionally adjusted for smoking status, physical activity and social class; model 3: additionally adjusted plasma vitamin C as marker of fruit and vegetable intake, tea and coffee intake; model 4: model 2, additionally adjusted for baseline health (self-reported diabetes mellitus, myocardial infarction, cerebrovascular accident), family history of myocardial infarction, use of anti-hypertensive or lipid-lowering drugs; model 5: model 3, additionally adjusted for baseline health (self-reported diabetes mellitus, myocardial infarction, cerebrovascular accident), family history of myocardial infarction, use of anti-hypertensive or lipid-lowering drugs. <sup>b</sup>Additionally adjusted for menopausal status and hormone replacement therapy. <sup>c</sup>Biomarker concentrations were adjusted by specific gravity.

The biomarkers used in this study were surrogate biomarkers<sup>29</sup>. They are used to rank participants according to intake, but it is not possible to use them to calculate the actual amount of flavan-3-ol consumed. In order to estimate the amount consumed by participants in the top decile of intake, we have therefore used data from 7-day food diaries and calculated flavan-3-ol intakes using not only mean food content, as is common practice, but the entire range of reported food content (Fig. 5). Participants in the top decile of flavan-3-ol intake consumed at least 146 mg/day when using minimum, and 618 mg/day when using maximum food content (260 mg/day when using mean food content data) of flavan-3-ols, and at least 4 mg/day (minimum food content) or 138 mg/day (maximum food content) of (-)-epicatechin (36 mg/day when using mean food content data). While these figures provide an initial estimation, further efforts are needed to obtain a more precise number if these data are expected to be used for the development of dietary recommendations of flavan-3-ols as bioactive.

**Strengths and limitations.** The EPIC-Norfolk cohort is ideally placed to investigate potential associations between flavan-3-ol intake and cardiovascular disease risk factors, not only because of its size, but also because it is set in a population with habitually high flavan-3-ol intake<sup>22</sup>. The main strength of the study is the use of robustly evaluated nutritional biomarkers to estimate flavan-3-ol<sup>27</sup> and (-)-epicatechin<sup>28</sup> intake and the use of 7-day-food diaries for dietary data. A limitation of the study is the reliance on a single spot-urine sample, as

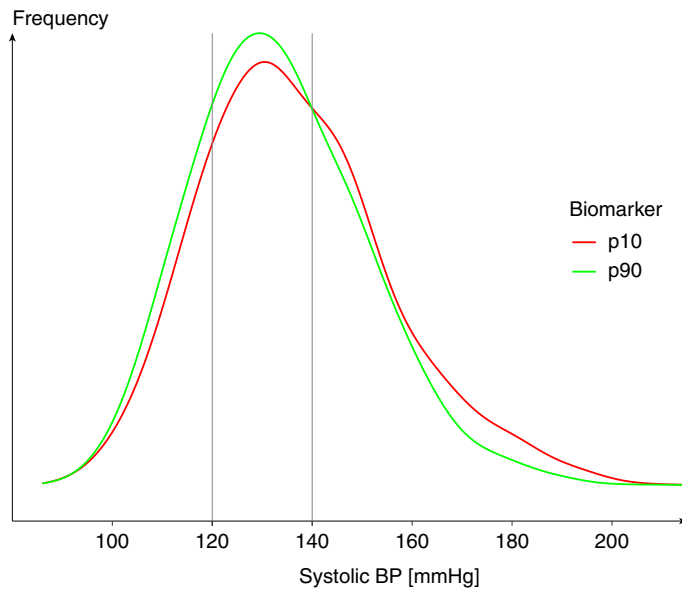


**Figure 3.** Subgroup- and sensitivity analysis, comparing estimated differences in systolic blood pressure between sex-specific bottom (p10) and top (p90) decile of biomarker-estimated flavan-3-ol intake. Models were adjusted by age, BMI, smoking status, physical activity and social class, and additionally for menopausal status and hormone-replacement therapy for women. <sup>†</sup>Hypertensive: systolic BP  $\geq 140$  mmHg; <sup>§</sup>normal weight: BMI  $\leq 25$  kg/m<sup>2</sup>; <sup>‡</sup>low risk: no baseline prevalence of diabetes or heart disease, no family history of heart disease, no anti-hypertensive or lipid-lowering drug use at baseline.

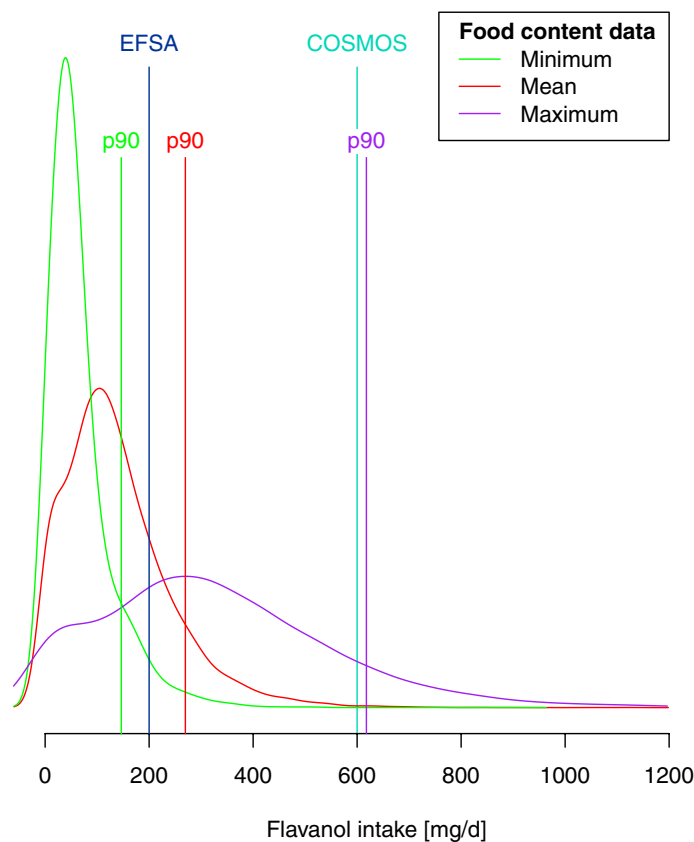
	Men (n = 11,592)		Women <sup>b</sup> (n = 14,026)	
	All-cause	CVD	All-cause	CVD
<b>Incidence</b>				
Model 0 <sup>a</sup>	–	1.04 (0.97; 1.11)	–	1.06 (0.99; 1.13)
Model 1	–	1.03 (0.96; 1.10)	–	1.03 (0.97; 1.10)
Model 2	–	1.02 (0.95; 1.09)	–	1.02 (0.95; 1.09)
Model 3	–	1.01 (0.95; 1.08)	–	1.01 (0.94; 1.07)
Model 4	–	1.04 (0.98; 1.11)	–	1.02 (0.96; 1.09)
Model 5	–	1.04 (0.98; 1.11)	–	1.01 (0.95; 1.08)
<b>Mortality<sup>c</sup></b>				
Model 0	0.99 (0.91; 1.08)	0.99 (0.86; 1.15)	1.00 (0.92; 1.09)	1.02 (0.87; 1.20)
Model 1	0.98 (0.90; 1.07)	0.99 (0.85; 1.14)	1.00 (0.91; 1.09)	1.01 (0.86; 1.19)
Model 2	0.97 (0.89; 1.05)	0.98 (0.84; 1.13)	0.97 (0.89; 1.06)	0.99 (0.85; 1.16)
Model 3	0.96 (0.88; 1.05)	0.97 (0.83; 1.12)	0.97 (0.88; 1.05)	1.00 (0.86; 1.18)
Model 4	0.98 (0.90; 1.07)	1.01 (0.88; 1.17)	0.97 (0.89; 1.06)	1.01 (0.86; 1.18)
Model 5	0.98 (0.90; 1.07)	1.01 (0.87; 1.17)	0.97 (0.89; 1.06)	1.03 (0.87; 1.20)

**Table 4.** Associations between gVLM<sub>B</sub> as biomarker of flavan-3-ol intake and CVD incidence and CVD and all-cause mortality. Results shown are estimated differences (HR, 95% CI) between low (10th percentile) and high (90th percentile) biomarker concentrations, using different statistical models. <sup>a</sup>Model 0: stratified by age-decade; model 1: additionally adjusted for BMI; model 2: additionally adjusted for smoking status, physical activity and social class; model 3: additionally adjusted plasma vitamin C as marker of fruit and vegetable intake, tea and coffee intake; model 4: model 2, additionally adjusted for baseline health (self-reported diabetes mellitus, myocardial infarction, cerebrovascular accident), family history of myocardial infarction, use of anti-hypertensive or lipid-lowering drugs; model 5: model 3, additionally adjusted for baseline health (self-reported diabetes mellitus, myocardial infarction, cerebrovascular accident), family history of myocardial infarction, use of anti-hypertensive or lipid-lowering drugs. <sup>b</sup>Additionally adjusted for menopausal status and hormone replacement therapy. <sup>c</sup>Biomarker concentrations were adjusted by specific gravity.





**Figure 4.** Distribution of systolic blood pressure of participants of EPIC Norfolk in the bottom (p10) and top (p90) decile of flavan-3-ol intake (estimated by  $gVLM_B$ ). Approximately 40% of participants in the bottom decile (p10) were hypertensive or pre-hypertensive (systolic blood pressure  $\geq 140$  mmHg), compared to 33% in the top decile.



**Figure 5.** Distribution of flavan-3-ol and (-)-epicatechin intake in EPIC Norfolk, estimated using 7-day food diaries and minimum (green), mean (red) and maximum (purple) food composition data<sup>20</sup>. The graph indicates the 90th percentile (p90) used as high intake in this study (146 mg/day when using minimum food content data, 270 mg/day when using mean food consumption data as is common practice and 618 mg/day when using maximum food composition data), as well as the amounts used for the EFSA approved health claim (200 mg/day)<sup>47</sup> and the COSMOS study (600 mg/day)<sup>11</sup>.

multiple samples would have provided a more representative estimation of habitual intake. The plasma half-life of gVLM, approximately 6 h, is sufficient to achieve steady-state like conditions with regular consumption<sup>27</sup>. As the main source of flavan-3-ol intake in a UK cohort is tea<sup>22</sup>, which is consumed regularly in a UK population, it can provide information on longer term intake. The half-life of SREM is considerably shorter, 2 h, and the biomarker therefore reflects mainly short-term intake. Further limitations are the largely cross-sectional nature and the inability to derive actual intake from surrogate biomarkers, only allowing the ranking of participants according to intake<sup>29</sup>.

**Conclusions.** This study demonstrates the importance of nutritional biomarkers to estimate intake of plant bioactives to investigate associations between intake and disease risk, as only biomarkers can provide reliable information on actual bioactive intake. It also raises the important question of the impact of the variability in food composition on nutritional research and dietary assessment. The results of our study show a significant and biomedically relevant inverse association between biomarkers of flavan-3-ol intake and blood pressure in a free-living general population with a wide range of flavan-3-ol intake from their habitual diet. The observed difference is approximately comparable to that of adherence with the Mediterranean Diet or moderate salt reduction, and likely to have a considerable impact on a population scale. In the context of an ageing population and increased prevalence of chronic diseases, these findings hold promise for the prevention of cardiovascular disease through dietary approaches.

## Methods

**Study population.** Between 1993 and 1997, 30,447 women and men aged between 40 and 75 years were recruited for the Norfolk cohort of the European Prospective Investigation into Cancer and Nutrition (EPIC) study, and 25,633 attended a health examination<sup>31</sup>. Health and lifestyle characteristics, including data on smoking, social class and family medical history, were assessed by questionnaire. Height and weight measurements were collected following a standardized protocol by trained research nurses. Physical activity, representing occupational and leisure activity, was assessed using a validated questionnaire<sup>48</sup>. Blood pressure was measured by using a non-invasive oscillometric blood pressure monitor (Acutorr; Datascope Medical, Huntingdon, UK; validated against sphygmomanometers every 6 months) after the participant had been seated in a comfortable environment for 5 min. The arm was horizontal and supported at the level of the mid-sternum; the mean of two readings was used for analysis. Non-fasting blood samples were taken by venepuncture and stored in serum tubes in liquid nitrogen. Serum levels of total cholesterol were measured on fresh samples with the RA 1000 autoanalyzer (Bayer Diagnostics, Basingstoke, UK). Plasma vitamin C was measured using a fluorometric assay as described previously<sup>49</sup>. Non-fasting spot urine samples were collected during the health examination and stored at  $-20^{\circ}\text{C}$  until analysis. Samples were collected throughout the day, and there were no consistent associations between collection time and biomarker concentration. Diet was assessed by 7-day diary (7DD), whereby the first day of the diary was completed as a 24-h recall (24HDR) with a trained interviewer and the remainder completed during subsequent days. Diary data were entered, checked and calculated using the in-house dietary assessment software DINER (Data into Nutrients for Epidemiological Research) and DINERMO<sup>50</sup>. Flavan-3-ol intake (the sum of epicatechin, catechin, epicatechin-3-O-gallate, catechin-3-O-gallate and proanthocyanidins) was estimated as described previously<sup>15</sup>; minimum and maximum estimated flavan-3-ol intake was estimated using the minimum and maximum food content data provided by Phenol Explorer and USDA databases<sup>20</sup>.

The study was approved by the Norwich Local Research Ethics Committee and all participants gave written, informed consent and all methods were carried out in accordance with relevant guidelines and regulations.

**Flavan-3-ol biomarker.** We have used two different biomarkers to estimate flavan-3-ol intake, one based on the flavan-3-ol-derived microbial metabolite 5-3',4'-dihydroxyphenyl- $\gamma$ -valerolactone (gVLM)<sup>27</sup>, and one based on structurally related (-)-epicatechin metabolites (SREM)<sup>28</sup>: gVLM<sub>B</sub> that includes the metabolites 5-(3',4'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'-O-glucuronide (gVL3G) and 5-(3',4'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'-sulphate (gVL3S), and SREM<sub>B</sub> that includes the metabolites (-)-epicatechin-3'-glucuronide (E3G), (-)-epicatechin-3'-sulfate (E3S) and 3'-O-methyl(-)-epicatechin-5-sulfate (3Me5S). gVLM<sub>B</sub> are specific for estimating the intake of flavan-3-ols in general, including ( $\pm$ )-epicatechin, ( $\pm$ )-catechin, ( $\pm$ )-epicatechin-3-O-gallate, ( $\pm$ )-catechin-3-O-gallate and procyanidins and excluding the flavan-3-ols galocatechin, epigallocatechin, galocatechin-3-O-gallate, epigallocatechin-3-O-gallate, theaflavins and thearubigins<sup>27</sup>. SREM<sub>B</sub> are specific for (-)-epicatechin intake<sup>28</sup>. Spot urine samples were collected during the baseline health examination and stored in glass bottles at  $20^{\circ}\text{C}$  until analysis. Stability analyses confirmed that biomarkers are stable under these conditions<sup>28</sup>. Samples were analysed in random order using the method described previously<sup>27,28</sup>, with automated sample preparation (Hamilton Star robot; Hamilton, Bonaduz, Switzerland). Briefly, 60  $\mu\text{L}$  spot urine sample and internal standard solutions (2.5  $\mu\text{M}$  <sup>13</sup>C<sub>2</sub>D<sub>2</sub>-5-(3',4'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'-sulphate, D<sub>2</sub>/D<sub>3</sub>-epicatechin-3'- $\beta$ -D-glucuronide, 50:50 mix) were combined, filtered (Impact Protein Precipitation filter plate, Phenomenex, Macclesfield, UK; centrifuged for 2 min at  $500 \times g$  at room temperature and stored at  $20^{\circ}\text{C}$  until analysis. Samples were then separated by liquid chromatography (Acquity, Waters, Elstree, UK) using a C18 column (Kinetex C18 200  $\times$  2.1 mm, 1.7  $\mu\text{m}$ , with 0.5  $\mu\text{m}$  Krudcatcher, Phenomenex, Macclesfield, UK) and detected by electrospray ionisation tandem mass spectrometry (Applied Biosystems API 4000, Sciex, Warrington, UK) in negative ion mode. The spray voltage was  $-4500\text{ V}$  and the source temperature was  $600^{\circ}\text{C}$ . Samples were quantified using calibration standards prepared in flavan-3-ol-metabolite free urine samples using the peak area ratio of analyte and internal standard. Each batch included two replicates of quality control samples with three different concentrations: low QC (0.3  $\mu\text{M}$ ), medium QC (2.5  $\mu\text{M}$  for SREM<sub>B</sub>, 25  $\mu\text{M}$  for gVLM<sub>B</sub>) and high QC (3.8  $\mu\text{M}$  for SREM<sub>B</sub>, 38  $\mu\text{M}$  for gVLM<sub>B</sub>) and usual acceptance criteria for each batch were that at

least one QC at each concentration and four out of the six QCs were within 15% of the theoretical concentration. The accuracy of the method was better than 5%, precision of the method (%CV) below 12% for all analytes. The results for 224 randomly inserted duplicate samples showed a high correlation ( $R^2 = 0.92$ ) with a mean difference of 1 (95% CI  $-0.04$ ;  $6.9$ )  $\mu\text{mol/L}$ . We did not observe any time-dependent change in method performance.

Concentrations below the lower limit of quantification (LLOQ,  $0.1 \mu\text{M}$ ) were used for the analysis to avoid the bias of substituting a range of values by a single value. Concentrations below the limit of detection were assigned a value of  $0.001 \mu\text{M}$ . Concentrations were adjusted by specific gravity for dilution as the endpoint of the analysis, systolic blood pressure, was strongly correlated with urinary creatinine<sup>51</sup>.

**Incident CVD events and mortality.** All participants were followed up for fatal and nonfatal CVD events, and the present study includes events until 31 March 2016, covering a median follow-up time of 19.5 (IQR 17.9; 20.9) years. Cause-specific hospital admission was determined via ENCORE (East Norfolk Commission Record, the hospital admissions database kept by the East Norfolk Health Commission)<sup>52</sup> with the individuals' unique National Health Service (NHS) number. All individuals were flagged by the UK Office of National Statistics (ONS) for death certification and trained nosologists coded death certificates according to the International Classification of Disease (ICD). The disease endpoints of this study was the first CVD event [defined as ICD 410–448 (ICD 9) or ICD I10–I79 (ICD 10)].

**Data analysis.** Data analyses were carried out using R 3.6<sup>53</sup>, using the packages *rms*<sup>54</sup> for regression analyses, *ggplot2*<sup>55</sup> and *gridExtra*<sup>56</sup> for the generation of graphics. Regression analyses were conducted using the *fit.mult.impute* function with either *ols* (cross-sectional analyses) or *cph* (prospective analyses) as regression function. We have used the Wald statistics calculated by the *rms anova* function to investigate the relationship between dependent and independent variables, and to test for linearity. *tableone*<sup>57</sup> was used to prepare tables. Unless indicated otherwise, results are shown with 95% confidence intervals.

**Descriptive statistics.** Descriptive characteristics of the study population were summarised using mean (standard deviation) for continuous variables and frequency (percentage) for categorical variables.

**Missing values.** Missing values (Supplemental Table 1) were assumed to be missing at random and were imputed using multiple imputation using the *aregImpute* algorithm<sup>54</sup>, which uses predictive mean matching with optional weighted probability sampling. We have used different imputed datasets: for the analysis of cross-sectional associations with the primary endpoint (blood pressure), for the analysis with cross-sectional associations with the secondary endpoints (blood lipids), and one for each prospective association with disease endpoints and mortality. In each imputation, we created 50 imputed, using restricted cubic splines (3 knots) and not assuming linearity, including all variables used in the final analysis. For prospective analyses, we have also included the Nelson-Aalen estimator<sup>58</sup>. There were no meaningful differences when comparing results obtained from the full data set with complete cases analyses in cross-sectional data (Supplementary Tables 3 and 4).

**Data transformation.** Biomarker data were positively skewed (log-normal distribution) and therefore log<sub>2</sub>-transformed data were used for all analyses. Restricted cubic splines (3 knots, outer quantiles 0.1 and 0.9; using the *rcs* function<sup>54</sup>) were used for all continuous variables unless indicated otherwise.

**Association between biomarker and self-reported intake.** Associations between biomarker-estimated flavan-3-ol intake and different food groups were investigated using multivariable regression analysis, using z-scores of self-reported food group intake ( $n = 96$ ) as independent variables. Models were adjusted for age, BMI, social class and energy intake, and stratified by sex.

**Cross-sectional analyses.** In cross-sectional analyses, stratified by sex, we investigated associations between the flavan-3-ol biomarker, (specific gravity adjusted concentration, log<sub>2</sub>-transformed), as an independent variable and systolic and diastolic blood pressure (mmHg) using multiple regression analyses. Statistical models were selected a priori based on likely confounders. Model 0 was adjusted for age (continuous; years); model 1 additionally for BMI (continuous;  $\text{kg/m}^2$ ); model 2: additionally for smoking status (categorical; never, ever, former), physical activity (categorical; inactive, moderately inactive, moderately active, active) and social class (categorical; unclassified, A, B, C1, C2, D, E); model 3: additionally for plasma vitamin C as marker of fruit and vegetable intake (continuous,  $\mu\text{M}$ ), tea and coffee intake (continuous, g/day, derived from 7-day diary); model 4: model 2, additionally adjusted for baseline health (self-reported diabetes mellitus, myocardial infarction, cerebrovascular accident), family history of myocardial infarction, use of anti-hypertensive or lipid-lowering drugs (all categorical; yes, no); model 5: model 3, additionally adjusted for baseline health (self-reported diabetes mellitus, myocardial infarction, cerebrovascular accident), family history of myocardial infarction, use of anti-hypertensive or lipid-lowering drugs (all categorical; yes, no). Analyses in women were additionally adjusted for menopausal status [categorical; pre-menopausal, peri-menopausal (2 categories), post-menopausal] and hormone replacement therapy (categorical; current, former, never).

**Association between biomarker and disease risk and mortality.** In prospective analyses, we investigated associations between the flavan-3-ol biomarker, (specific gravity adjusted concentration, log<sub>2</sub>-transformed), as an independent variable and disease risk (CVD, IHD, MI and Stroke), and mortality using Cox regression analyses. The proportional hazard assumptions has been tested using the *cox.zph* function and was

met. Statistical models were selected a priori based on likely confounders. All models were stratified by age (as decade) at baseline, but age was not included as a covariable. Otherwise, models were the same as described for cross-sectional analyses.

**Sensitivity and complete cases analyses.** Post-hoc sensitivity analyses were conducted by restricting the study population to those without any self-reported disease or disease risk at baseline and post-menopausal women. The same analyses as described above were also conducted on a subsample of the study population for whom all data were available (complete cases analysis).

Received: 21 April 2020; Accepted: 6 October 2020

Published online: 21 October 2020

## References

1. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: A systematic analysis for the global burden of disease study 2013. *Lancet* **385**, 117–171 (2015).
2. Van Horn, L. *et al.* Recommended dietary pattern to achieve adherence to the American Heart Association/American College of Cardiology (AHA/ACC) guidelines: A scientific statement from the American Heart Association. *Circulation* **134**, e505–e529 (2016).
3. Yetley, E. A. *et al.* Options for basing dietary reference intakes (dris) on chronic disease endpoints: Report from a joint us-/canadian-sponsored working group. *Am. J. Clin. Nutr.* **105**, 249S–285S (2017).
4. Schroeter, H. *et al.* Recommending flavanols and procyanidins for cardiovascular health: Current knowledge and future needs. *Mol. Aspects Med.* **31**, 546–557 (2010).
5. Lupton, J. R. *et al.* Exploring the benefits and challenges of establishing a dri-like process for bioactives. *Eur. J. Nutr.* **53**(Suppl 1), 1–9 (2014).
6. Ottaviani, J. I., Heiss, C., Spencer, J. P. E., Kelm, M. & Schroeter, H. Recommending flavanols and procyanidins for cardiovascular health: Revisited. *Mol. Aspects Med.* **61**, 63–75 (2018).
7. Ried, K., Fakler, P. & Stocks, N. P. Effect of cocoa on blood pressure. *Cochrane Database Syst. Rev.* **4**, CD008893 (2017).
8. Schroeter, H. *et al.* (-)-epicatechin mediates beneficial effects of flavanol-rich cocoa on vascular function in humans. *Proc. Natl. Acad. Sci.* **103**, 1024–1029 (2006). <https://www.pnas.org/content/103/4/1024.full.pdf>.
9. van Praag, H. *et al.* Plant-derived flavanol (-)-epicatechin enhances angiogenesis and retention of spatial memory in mice. *J. Neurosci.* **27**, 5869–5878 (2007).
10. Sansone, R. *et al.* Cocoa flavanol intake improves endothelial function and framingham risk score in healthy men and women: A randomised, controlled, double-masked trial: the flaviola health study. *Br. J. Nutr.* **114**, 1246–1255 (2015).
11. Rautiainen, S., Sesso, H. D. & Manson, J. E. Large-scale randomized clinical trials of bioactives and nutrients in relation to human health and disease prevention - lessons from the vital and cosmos trials. *Mol. Aspects Med.* **61**, 12–17 (2018).
12. Mink, P. J. *et al.* Flavonoid intake and cardiovascular disease mortality: A prospective study in postmenopausal women. *Am. J. Clin. Nutr.* **85**, 895–909 (2007).
13. McCullough, M. L. *et al.* Flavonoid intake and cardiovascular disease mortality in a prospective cohort of us adults. *Am. J. Clin. Nutr.* **95**, 454–464 (2012).
14. Cassidy, A. *et al.* Habitual intake of flavonoid subclasses and incident hypertension in adults. *Am. J. Clin. Nutr.* **93**, 338–347 (2011).
15. Vogiatzoglou, A. *et al.* Associations between flavan-3-ol intake and risk of CVD in the norfolk cohort of the European prospective investigation into cancer (EPIC-Norfolk). *Free Radic. Biol. Med.* **84**, 1–10 (2015).
16. Bondonno, N. P. *et al.* Flavonoid intake is associated with lower mortality in the danish diet cancer and health cohort. *Nat. Commun.* **10**, 3651–10 (2019).
17. Subar, A. F. *et al.* Addressing current criticism regarding the value of self-report dietary data. *J. Nutr.* **145**, 2639–2645 (2015).
18. Park, Y. *et al.* Comparison of self-reported dietary intakes from the automated self-administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers. *Am. J. Clin. Nutr.* **107**, 80–93 (2018).
19. Kuhnle, G. G. *et al.* Association between sucrose intake and risk of overweight and obesity in a prospective sub-cohort of the european prospective investigation into cancer in norfolk (epic-norfolk). *Public Health Nutr.* **18**, 2815–2824 (2015).
20. Rothwell, J. A. *et al.* Phenol-explorer 3.0: A major update of the phenol-explorer database to incorporate data on the effects of food processing on polyphenol content. *Database* **2013**, bat070 (2013).
21. Haytowitz, D. B., Bhagwat, S. & Holden, J. M. Sources of variability in the flavonoid content of foods. *Proc. Food Sci.* **2**, 46–51 (2013).
22. Vogiatzoglou, A. *et al.* Assessment of the dietary intake of total flavan-3-ols, monomeric flavan-3-ols, proanthocyanidins and theaflavins in the european union. *Br. J. Nutr.* **111**, 1463–1473 (2014).
23. Kuhnle, G. G. Nutrition epidemiology of flavan-3-ols: The known unknowns. *Mol. Aspects Med.* **61**, 2–11 (2017).
24. Kaaks, R., Riboli, E. & Sinha, R. Biochemical markers of dietary intake. *IARC Sci. Publ. No.* **84**(84), 103–126 (1997).
25. Jenab, M., Slimani, N., Bictash, M., Ferrari, P. & Bingham, S. A. Biomarkers in nutritional epidemiology: Applications, needs and new horizons. *Hum. Genet.* **125**, 507–525 (2009).
26. Kuhnle, G. G. C. Nutritional biomarkers for objective dietary assessment. *J. Sci. Food Agric.* **92**, 1145–1149 (2012).
27. Ottaviani, J. I. *et al.* Evaluation at scale of microbiome-derived metabolites as biomarker of flavan-3-ol intake in epidemiological studies. *Sci. Rep.* **8**, 9859 (2018).
28. Ottaviani, J. I. *et al.* Evaluation of (-)-epicatechin metabolites as recovery biomarker of dietary flavan-3-ol intake. *Sci. Rep.* **9**, 13108–10 (2019).
29. Keogh, R. H., White, I. R. & Bingham, S. A. Using surrogate biomarkers to improve measurement error models in nutritional epidemiology. *Stat. Med.* **32**, 3838–3861 (2013).
30. Harding, A.-H. *et al.* Plasma vitamin C level, fruit and vegetable consumption, and the risk of new-onset type 2 diabetes mellitus: The European prospective investigation of cancer-norfolk prospective study. *Arch. Intern. Med.* **168**, 1493–1499 (2008).
31. Day, N. *et al.* EPIC-Norfolk: Study design and characteristics of the cohort. European prospective investigation of cancer. *Br. J. Cancer* **80**(Suppl 1), 95–103 (1999).
32. Wilkinson, B. G. & Perring, M. A. Variation in mineral composition of cox's orange pippin apples. *J. Sci. Food Agric.* **12**, 74–80. <https://doi.org/10.1002/jsfa.2740120114> (1961).
33. Schroeter, H. *et al.* (-)-epicatechin mediates beneficial effects of flavanol-rich cocoa on vascular function in humans. *Proc. Nat. Acad. Sci. U.S.A.* **103**, 1024–1029 (2006).

34. Rodriguez-Mateos, A. *et al.* Assessing the respective contributions of dietary flavanol monomers and procyanidins in mediating cardiovascular effects in humans: randomized, controlled, double-masked intervention trial. *Am. J. Clin. Nutr.* **31**, 546–1237 (2018).
35. Heiss, C. *et al.* Improvement of endothelial function with dietary flavanols is associated with mobilization of circulating angiogenic cells in patients with coronary artery disease. *J. Am. Coll. Cardiol.* **56**, 218–224 (2010).
36. Ruskovska, T. *et al.* Systematic bioinformatic analysis of nutrigenomic data of flavanols in cell models of cardiometabolic disease. *Food Funct.* **11**, 5040–5064 (2020).
37. Milenkovic, D. *et al.* (-)-Epicatechin metabolites promote vascular health through epigenetic reprogramming of endothelial-immune cell signaling and reversing systemic low-grade inflammation. *Biochem. Pharmacol.* **173**, 113699 (2020) (**Natural Products and the Hallmarks of Chronic Diseases**).
38. Samavat, H. *et al.* Effects of green tea catechin extract on serum lipids in postmenopausal women: A randomized, placebo-controlled clinical trial. *Am. J. Clin. Nutr.* **104**, 1671–1682 (2016).
39. Hippisley-Cox, J., Coupland, C. & Brindle, P. Development and validation of qrisk3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ (Clin. Res. Ed.)* **357**, j2099 (2017).
40. Toledo, E. *et al.* Effect of the mediterranean diet on blood pressure in the predimed trial: Results from a randomized controlled trial. *BMC Med.* **11**, 207 (2013).
41. Sacks, F. M. *et al.* Soy protein, isoflavones, and cardiovascular health: An American Heart Association science advisory for professionals from the nutrition committee. *Circulation* **113**, 1034–1044 (2006).
42. Erlinger, T. P., Vollmer, W. M., Svetkey, L. P. & Appel, L. J. The potential impact of nonpharmacologic population-wide blood pressure reduction on coronary heart disease events: pronounced benefits in african-americans and hypertensives. *Prev. Med.* **37**, 327–333 (2003).
43. Whelton, P. K. *et al.* Effects of oral potassium on blood pressure meta-analysis of randomized controlled clinical trials. *JAMA J. Am. Med. Assoc.* **277**, 1624–1632 (1997).
44. Law, M. R., Frost, C. D. & Wald, N. J. By how much does dietary salt reduction lower blood pressure? I-Analysis of observational data among populations. *BMJ (Clin. Res. Ed.)* **302**, 811–815 (1991).
45. Sacks, F. M. *et al.* Effects on blood pressure of reduced dietary sodium and the dietary approaches to stop hypertension (dash) diet. Dash-sodium collaborative research group. *N. Engl. J. Med.* **344**, 3–10 (2001).
46. Stamler, J. The intersalt study: Background, methods, findings, and implications. *Am. J. Clin. Nutr.* **65**, 626S–642S (1997).
47. EFSA Panel on Dietetic Products, Nutrition, and Allergies (NDA). Scientific opinion on the modification of the authorisation of a health claim related to cocoa flavanols and maintenance of normal endothelium-dependent vasodilation pursuant to article 13(5) of regulation (EC) no 1924/2006 following a request in accordance with article 19 of regulation (ec) no 1924/2006. *EFSA J.* **12**, 3654 (2014).
48. Wareham, N. J. *et al.* Validity and repeatability of the EPIC-Norfolk physical activity questionnaire. *Int. J. Epidemiol.* **31**, 168–174 (2002).
49. Khaw, K.-T. *et al.* Relation between plasma ascorbic acid and mortality in men and women in EPIC-Norfolk prospective study: A prospective population study. *Lancet* **357**, 657–663 (2001).
50. Lentjes, M. *et al.* Dietary intake measurement using 7 d diet diaries in british men and women in the european prospective investigation into cancer-norfolk study: a focus on methodological issues. *Br. J. Nutr.* **111**, 516–526 (2014).
51. Bingham, S. A. *et al.* Epidemiologic assessment of sugars consumption using biomarkers: Comparisons of obese and nonobese individuals in the european prospective investigation of cancer norfolk. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Research, cosponsored by the American Society of Preventive Oncology* **16**, 1651–1654 (2007).
52. Sinha, S., Myint, P. K., Luben, R. N. & Khaw, K.-T. Accuracy of death certification and hospital record linkage for identification of incident stroke. *BMC Med. Res. Methodol.* **8**, 74 (2008).
53. Team, R. C. R: A language and environment for statistical computing. *J. Article* (2018).
54. Harrell Jr, F. E. RMS: Regression modeling strategies. *Journal Article* (2017).
55. Wickham, H. *ggplot2: Elegant graphics for data analysis* (Springer, New York, 2009).
56. Auguie, B. gridextra: Miscellaneous functions for grid graphics. *Report* (2017).
57. Yoshida, K. & Bohn, J. tableone: Create table to describe baseline characteristics. *Report* (2015).
58. White, I. R. & Royston, P. Imputing missing covariate values for the cox model. *Stat. Med.* **28**, 1982–1998 (2009).

## Acknowledgements

We thank all EPIC-Norfolk study participants and staff for their contribution to the study.

## Author contributions

J.I.O., H.S. and G.G.C.K. designed the research. J.I.O., R.F., P.B.G., D.H.M., A.T., A.W. and G.G.C.K. developed and validated the analytical methods. A.B., N.G., D.L., P.B.G., D.H.M., A.T., A.W., R.L. and G.G.C.K. analysed and processed samples of the prospective cohort study. K.T.K. and N.G.F. are PIs of EPIC Norfolk, provided intellectual input and edited the manuscript. J.I.O., H.S. and G.G.C.K. had primary responsibility for the final content. All authors read and approved the final manuscript.

## Funding

Funding was provided by Medical Research Council (Grant no. MC\_UU\_12015/5), National Institute for Health Research (Grant no. IS-BRC-1215-20014) and Mars, Inc.

## Competing interests

H.S. and J.I.O. are employed by Mars, Inc., a company engaged in flavanol research and flavanol-related commercial activities. G.G.C.K. has received an unrestricted research grant from Mars, Inc. None of the other authors has a conflict of interest to declare.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-74863-7>.

**Correspondence** and requests for materials should be addressed to G.G.C.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020