Fruit and vegetable consumption and all-cause mortality: a dose-response analysis

Andrea Bellavia, Susanna C Larsson, Matteo Bottai, Alicja Wolk, and Nicola Orsini

ABSTRACT
Background: The association between fruit and vegetable (FV) consumption and overall mortality has seldom been investigated in large cohort studies. Findings from the few available studies are inconsistent.

Objective: The objective was to examine the dose-response relation between FV consumption and mortality, in terms of both time and rate, in a large prospective cohort of Swedish men and women.

Design: FV consumption was assessed through a self-administrated questionnaire in a population-based cohort of 71,706 participants (38,221 men and 33,485 women) aged 45–83 y. We performed a dose-response analysis to evaluate 10th survival percentile differences (PDs) by using Laplace regression and estimated HRs by using Cox regression.

Results: During 13 y of follow-up, 11,439 deaths (6803 men and 4636 women) occurred in the cohort. In comparison with 5 servings FV/d, a lower consumption was progressively associated with shorter survival and higher mortality rates. Those who never consumed FV lived 3 y shorter (PD: −37 mo; 95% CI: −58, −16 mo) and had a 53% higher mortality rate (HR: 1.53; 95% CI: 1.19, 1.99) than did those who consumed 5 servings FV/d. Consideration of fruit and vegetables separately showed that those who never consumed fruit lived 19 mo shorter (PD: −19 mo; 95% CI: −29, −10 mo) than did those who ate 1 fruit/d. Participants who consumed 3 vegetables/d lived 32 mo longer than did those who never consumed vegetables (PD: 32 mo; 96% CI: 13, 51 mo).

Conclusion: FV consumption <5 servings/d is associated with progressively shorter survival and higher mortality rates. The Swedish Mammography Cohort and the Cohort of Swedish Men were registered at clinicaltrials.gov as NCT01127698 and NCT01127711, respectively. Am J Clin Nutr doi: 10.3945/ajcn.112.056119.

INTRODUCTION
Consumption of ≥5 servings of fruit and vegetables (FVs) per day is suggested by dietary and food recommendations (1). Major focus has been given to the effect of FV consumption on specific chronic diseases such as cardiovascular disease (CVD) (2, 3) and cancer (4). Few cohort studies have investigated FV consumption in relation to risk of overall mortality as a primary outcome (5–9), and most of the available information comes from smaller studies that examined total serum carotenoids as a marker of FV intake (10–14). The association was mainly evaluated by categorizing the main variable in quintiles or tertiles of daily consumption (6–9, 11–14), despite the recognized limitations of the use of categories (15, 16). Findings from those studies are inconsistent. To the best of our knowledge, no previous studies have evaluated FV consumption as a continuous variable to investigate the shape of the association between FV consumption and mortality in a population-based cohort study.

Therefore, our aim was to examine the dose-response relation between FV consumption and mortality, in terms of both time to death and mortality rate, in a large cohort of Swedish men and women.

SUBJECTS AND METHODS
Study population
Participants from the population-based Cohort of Swedish Men and the Swedish Mammography Cohort were combined in this study. Briefly, participants in the Cohort of Swedish Men were recruited in 1997–1998, when all men who were 45–79 y of age and resided in Västmanland and Örebro counties (central Sweden) received an invitation to participate in the study. A self-administrated questionnaire was used to collect information on diet, alcohol consumption, education, body weight, height, physical activity, smoking habits, and other lifestyle factors. A total of 48,850 men returned the questionnaire. The population-based Swedish Mammography Cohort was established between 1987 and 1990 when women born between 1914 and 1948 and residing in Västmanland and Uppsala counties (central Sweden) were recruited. Participants completed a questionnaire with questions regarding diet, alcohol consumption, education, body weight, and height. In the late fall of 1997, women who were still alive and residing in the study area received a second questionnaire that was expanded to include information regarding smoking status, physical activity, and other lifestyle factors. A total of 39,227 women returned this second ques-

1From the Units of Nutritional Epidemiology (AB, SCL, AW, and NO) and Biostatistics (AB, MB, and NO), Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.
2Supported in part by a Young Scholar Award from the Karolinska Institutes’s Strategic Program in Epidemiology, the Swedish Medical Society (SLS-250271), and the Swedish Research Council.
3Address correspondence to N Orsini, Unit of Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, PO Box, SE-171 77, Stockholm, Sweden. E-mail: nicola.orsini@ki.se.
4Abbreviations used: CVD, cardiovascular disease; FV, fruit and vegetable; PD, percentile difference.
Received December 3, 2012. Accepted for publication May 6, 2013. doi: 10.3945/ajcn.112.056119.
tionnaire and were considered in the current analysis. This study was approved by the Regional Research Ethics Board at Karolinska Institutet, and all participants gave their informed consent.

For the analyses, we excluded participants who reported incorrect or missing personal numbers ($n = 540$), those who died before the start of follow-up ($n = 97$), and those who had diabetes ($n = 3534$) or a history of CVD ($n = 6994$) or cancer ($n = 4390$). We also excluded participants who reported an implausible value for total energy intake (3 SDs from the log transformed mean energy intake, $n = 676$) and those with an unlikely high value of FV consumption (>20 servings/d) or missing data on all variables about FV ($n = 140$). After these exclusions, a total of 71,706 participants (38,221 men and 33,485 women) were included in the current analysis.

**FV consumption assessment**

Usual dietary intake over the previous year was assessed through a self-administered 96-item food-frequency questionnaire. In our validation study, the Spearman correlation coefficients between the average of four 1-wk dietary records and the dietary questionnaire ranged from 0.4 to 0.7 for individual FV items (A Wolk, unpublished data, 1992). Information on FV consumption was collected by using 14 questions on vegetables (carrot, beetroot, lettuce, cabbage, cauliflower, broccoli, tomato, pepper, spinach, peas, onion, garlic, pea soup, and other vegetables), 5 on fruit (orange, apple, banana, berry, and other fruit), and 1 on orange juice. Participants were asked to indicate how often on average in the previous year they had consumed each food. Our main variables for FV, fruit, and vegetable consumption were calculated as the average servings per day and were obtained by converting the questionnaire responses to average daily intake of each item and adding the intake of all items. A total of 46% of the participants provided information on all 20 questions about FV and almost 80% of the participants reported <2 missing values. When aggregating items, we assumed that missing values for an individual food meant no intake for that particular item (17).

**Case ascertainment and follow-up**

From 1 January 1998 through 31 December 2010, during 13 y of follow-up, we documented 11,439 deaths in the cohort (6803 men and 4636 women). Information on death was ascertained from the Swedish Register of Death Causes at the National Board of Health and Welfare. It has been estimated that 93% of all deaths in Sweden are reported within 10 d, and 100% are reported within 30 d (18).

**Statistical analysis**

We used multivariable Laplace regression to model percentiles of survival (19, 20). In our cohort, 16% of participants died during the follow-up period. Therefore, we focused our analysis on the 10th percentile of survival, expressed in months, that is the time by which 10% of participants in the current study has died. The evaluation of other percentiles between 1 and 16 provided similar results. The main measure of exposure-disease association was defined as the 10th percentile difference (PD). Mortality rates were modeled by using multivariable Cox regression, and HRs are reported. Proportional hazards assumption was checked by calculating Schoenfeld’s residuals, regressed against survival time, and tested for a nonzero slope. We found no evidence of departure from the assumption.

Our multivariable analyses were adjusted for sex, age at baseline (<50, 50–54, 55–59, 60–64, 65–69, 70–74, or ≥75 y), BMI (in kg/m$^2$; <25, 25–29, or ≥30), total physical activity (continuous variable; MET-h/d), smoking status and pack-years of smoking (current ≥40, current 20–39, current <20, former ≥40, former 20–39, former <20, or never), alcohol consumption (never drinker or <5, 5–10, 10.1–20, or >20 g/d), education level (primary school, high school, or university), and total energy intake (continuous variable; kcal/d).

We evaluated age-standardized characteristics of the study population by categories of FV consumption classifying our main variable into 4 categories (≤2.0, 2.1–4.0, 4.1–6.0, or >6 servings/d). Age-adjusted $P$-trend values of potential confounders across exposures categories were obtained from a linear regression for continuous variables, logistic regression for binary variables, and ordinal logistic regression for categorical variables. To evaluate the dose-response relation between FV consumption and mortality in terms of survival time and mortality rate, we flexibly modeled our continuous exposure with the use of right-restricted cubic splines with 3 knots of the distribution (at 3, 5, and 8 servings of FV/d). The shape of the dose-response relation was fairly insensitive to the location of the knots (21). We evaluated differences in the 10th survival percentile and mortality rates using the dose of 5 servings/d as referent. Linearity was evaluated by testing the null hypothesis that the coefficients of the unrestricted spline transformations are jointly equal to zero (22).

We also evaluated the dose-response relation for fruit and vegetable consumption as separate exposures in a single model. Because the distributions of fruit and vegetable consumption in our cohort were substantially different, we estimated differences in the 10th survival percentiles with the use of different reference values (1 serving fruit/d and 3 servings vegetables/d).

To reduce the potential effects of undiagnosed diseases on diet (participants with a known history of CVD, cancer, and diabetes were already excluded at baseline), we performed a sensitivity analysis to evaluate the association between FV consumption and mortality excluding participants who died during the first 3 y of follow-up ($n = 1243$). A second sensitivity analysis was performed by further adjusting our models for the nonrecommended food score as previously described (23). In brief, the score was built including information on 16 food items (3 red meat products, 5 processed meat products, 3 high-fat dairy products, white bread, sweets, chips/fries, mayonnaise, and ice cream). Consumption of any of these nonrecommended products ≥3 times/wk was assigned 1 point; the maximum score was 16 points (23). In our last sensitivity analysis, we evaluated whether there was any difference excluding information on orange juice.

We next performed a secondary analysis to investigate the potential interaction of FV with sex, smoking status (current compared with former/never), BMI (continuous), and education level (high school/ university compared with primary) in predicting time to death. Statistical interactions were assessed by testing the product terms in the model with a Wald test. Statistical analyses were performed with Stata (version 12; StataCorp). All
RESULTS

The characteristics of the study population by categories of FV consumption are shown in Table 1. On average, women tended to consume more FV than did men. Participants with a low FV intake were more likely to be current smokers, to have a lower educational level, and to have a higher consumption of non-recommended foods. An increase in FV consumption corresponded to a higher total energy intake. Age, BMI, physical activity, and alcohol consumption were similar, overall, across categories of FV consumption. The overall 10th survival percentile was 116 mo (95% CI: 114, 118 mo), which meant that 90% of the cohort was still alive after 9.6 y of follow-up.

The dose-response association between FV consumption and mortality is shown in Figure 1. We flexibly modeled the association by using splines and estimating 10th survival PDs and HRs. We observed strong evidence of departure from linearity (P < 0.001). The point and interval estimates of the dose-response analysis at specific levels of the distribution of FV consumption are shown in Table 2. Compared with an FV consumption of 5 servings/d, lower levels of consumption were progressively associated with shorter survival up to 3 y for those who never consumed FV daily (PD: −37 mo; 95% CI: −58, −16 mo). Consumption of FV >5 servings/d was not associated with a significant improvement in survival. The mortality rate for those who did not consume FVs was 53% higher (HR: 1.53; 95% CI: 1.19, 1.99) than the rate for those who consumed the recommended dose of 5 servings/d.

We next considered the consumption of fruit and vegetables separately in a mutually adjusted model. The 10th survival PDs, according to levels of fruit consumption and vegetables consumption, are presented in Figure 2. We found strong evidence of nonlinearity for consumption of both fruit and vegetables (P < 0.001). Compared with those who consumed 1 serving fruit/d, those who never consumed fruit lived 19 mo less (PD: −19 mo; 95% CI: −29, −10). Consumption of >1 serving fruit/d was not associated with a significant increase in survival compared with consumption of 1 serving/d. Daily vegetable consumption was associated with longer survival up to 3 servings/d. Participants who consumed 3 servings vegetables/d lived 32 mo longer than did those who never used to eat vegetables (PD: 32 mo; 95% CI: 13, 51 mo).

In our first sensitivity analysis, we excluded participants who died during the first 3 y of follow-up. The magnitude of the associations was slightly attenuated, but we observed no changes in the statistical significance or in the shape of the dose-response relation when comparing 0–5 servings FV/d (PD: −23 mo; 95% CI: −40, −5 mo; HR: 1.45; 95% CI: 1.09, 1.91). In the second sensitivity analysis, we further adjusted for the nonrecommended food score and observed negligible changes in the results when comparing 0–5 servings FV/d (PD: −36 mo; 95% CI: −63, −9 mo and HR: 1.54; 95% CI: 1.19, 2.00). Finally, no changes were observed when we omitted orange juice from total FV or fruit consumption.

We next evaluated possible interactions between FV and sex, smoking status, BMI, and education level in predicting mortality. The shape of the dose-response relation between FV and the 10th survival percentile did not substantially change according to sex (P-interaction = 0.31), smoking status

### Table 1

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Categories of daily fruit and vegetable consumption (median)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≤2 (1.5)</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------</td>
</tr>
<tr>
<td>No. of subjects</td>
<td>11,922</td>
</tr>
<tr>
<td>Female (%)</td>
<td>29</td>
</tr>
<tr>
<td>Mean age at baseline (y)</td>
<td>61.5</td>
</tr>
<tr>
<td>Mean BMI (kg/m²)</td>
<td>25.6</td>
</tr>
<tr>
<td>Total physical activity (MET-h/d)</td>
<td>41.8</td>
</tr>
<tr>
<td>Smoking status (%)</td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>34</td>
</tr>
<tr>
<td>Former</td>
<td>29</td>
</tr>
<tr>
<td>Never</td>
<td>37</td>
</tr>
<tr>
<td>Alcohol consumption (%)</td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>84</td>
</tr>
<tr>
<td>Former</td>
<td>5</td>
</tr>
<tr>
<td>Never</td>
<td>11</td>
</tr>
<tr>
<td>High consumption of nonrecommended food (%)</td>
<td>70</td>
</tr>
<tr>
<td>High school/university (%)</td>
<td>17</td>
</tr>
<tr>
<td>Energy (kcal/d)</td>
<td>2058</td>
</tr>
</tbody>
</table>

1 All variables except age were directly standardized to the age distribution of the entire study cohort (n = 71,706).
2 Age-adjusted P-trend values were obtained from a linear regression for continuous variables, logistic regression for binary variables, and ordinal logistic regression for categorical variables.
3 MET, metabolic equivalent task.
4 Proportion of participants with a nonrecommended food score >8 at a range between 0 and 16.
when comparing the first quintile of consumption (median: 0.9 servings/d) with the fifth (median: 4.9 servings/d) (8). A strong risk reduction for those who consume FV is supported by another recent small study (n = 713) on total serum carotenoids reporting an HR of 0.50 for women in the highest compared with the lowest tertile (14). On the other hand, other studies have found that this decrease in mortality risk was not substantial (6, 7). One of the major studies on FV consumption and mortality evaluated the association in a large cohort of >100,000 participants and found a nonsignificant HR of 0.95 when comparing the fifth (median: 9.2 servings/d) with the first (median: 2.6 servings/d) quintile (7). Another relatively large study (n = 16,000) observed a nonsignificant risk reduction when comparing the second to fifth quintiles with the first quintile (median: 1.5 servings/d) (6). A possible explanation for the discrepancy in the previous findings is the choice to perform analyses only by categories of the quantitative exposure obtained by using study-specific cutoffs (quintiles or tertiles). This approach had some limitations that have been widely identified (15, 16, 24). The major limitations of a categorical approach are the assumption of a step dose-response function, the subjective choice of cutoff, the loss of statistical power, and the loss of within-category information. Categorization by necessity leads to pooling groups with different risks so that any difference between individuals in the same category cannot be detected. We observed a strong increase in mortality among those with low levels of consumption. This decrease could not be observed if all participants with low consumption were considered together in the same category. Moreover, trying to characterize the shape of the dose-response association from a categorical approach may not be straightforward in the presence of strong nonlinearity, as we observed in our data. Although one study on serum carotenoid

### Table 2

<table>
<thead>
<tr>
<th>Fruit and Vegetable Consumption</th>
<th>10th Percentile Differences (months)</th>
<th>95% CI</th>
<th>HR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
<td>1.53</td>
<td>1.19, 1.99</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>25</td>
<td>1.37</td>
<td>1.20, 1.56</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>1.26</td>
<td>1.17, 1.37</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>1.16</td>
<td>1.09, 1.24</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1.11</td>
<td>1.05, 1.17</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1.05</td>
<td>1.02, 1.09</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.97</td>
<td>0.93, 1.01</td>
<td></td>
</tr>
</tbody>
</table>

1 Tabular presentation of estimates for specific values of fruit and vegetable consumption from the spline dose-response model. The value of 5 servings/d was used as the referent.
2 Percentile differences (PD) in the 10th survival percentile and HRs were adjusted for baseline age (<50, 50–54, 55–59, 60–64, 65–69, 70–74, or >75 y), sex (male or female), BMI (in kg/m²; <25, 25–29, or ≥30), total physical activity (MET-h/d; continuous variable), smoking status (current or not), and categorical alcohol consumption (never, <5, 5–10, 10.1–20, or >20 g/d), and educational level (primary school, high school, or university).
intake has already suggested that the shape of the relation could be nonlinear (13), a linear trend was often assumed and positively tested (6–9).

To the best of our knowledge this was the first cohort study that evaluated the dose-response relation between FV consumption and mortality with the use of flexible tools such as splines. The strengths of this study were the population-based and prospective design, large sample size, completeness of ascertainment of deaths through the National Register, and detailed information on diet. The small fraction of missing data reported on single items, which were treated with the zero-consumption approach, is unlikely to represent a source of bias for the observed findings (17). Another major strength of this work was that we characterized the association both in terms of time and in terms of rate—an approach that makes it easier to interpret results and communicate them to the general public. Information on time to event is obtained by the estimation of survival percentiles, which express the time period within which a specified proportion of people die. Laplace regression, the main statistical model that we used for analyses, directly estimates differences in survival percentiles according to levels of the exposure (19, 20). This approach provided many advantages, such as modeling continuous exposures, adjusting for potential confounders, and assessing interactions in predicting survival. The estimation of differences in survival percentiles (the regression coefficients obtained from Laplace regression) provides an intuitive measure of the observed association between the exposure of interest and mortality directly in the unit of the time scale (eg. months, years). In our closed cohort, all censored observations occurred at the end of the study period, and estimable survival percentiles depended on the length of follow-up. Inference on higher percentiles, for example median survival time, would require extrapolation beyond the range of observed follow-up time. No other statistical methods, Cox regression included, could overcome this limitation, which is inherent in this type of data.

The main limitation of this study was that information on FV consumption was self-reported, which can lead to a potential misclassification of the exposure. Classification errors in our prospective study, however, were nondifferential with respect to the occurrence of death and most likely led to an attenuation of the results.

In summary, the findings from this prospective study indicate that FV consumption <5 servings/d is associated with progressively shorter survival and higher mortality rates. FV consumption >5 servings/d does not provide considerable added benefits with respect to survival.

The authors' responsibilities were as follows—AW: was responsible for data collection; AB: was responsible for the statistical analyses; AB, SCL, MB, AW, and NO: were responsible for the interpretation of the results; AB: drafted the manuscript under the supervision of NO; and SCL, MB, and AW: reviewed and revised the manuscript. All authors reviewed and approved the final manuscript and participated in the study design and in the writing of the manuscript. None of the authors had any personal or financial conflicts of interest. The funders had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

REFERENCES


