Assessing preventability of postmenopausal breast cancer by lifestyle risk factors: influence by method, data source, and timing of exposure measurement

You Wu (youwu@tsinghua.edu.cn)
Tsinghua University

Hanseul Kim
Massachusetts General Hospital

Kai Wang
Harvard University T H Chan School of Public Health

Mingyang Song
Harvard University T H Chan School of Public Health

Molin Wang
Harvard University T H Chan School of Public Health

Rulla Tamimi
Weill Cornell Medicine

Heather Eliassen
Harvard University T H Chan School of Public Health

Stephanie Smith-Warner
Harvard University T H Chan School of Public Health

Walter Willett
Harvard University T H Chan School of Public Health

Edward Giovannucci
Harvard University T H Chan School of Public Health

Research Article

Keywords: population attributable risk, lifestyle, breast cancer, cancer preventability, repeated measurements

Posted Date: July 29th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1761955/v1

Abstract

Background

Population attributable risk (PAR%) reflects the preventable fraction of disease. However, PAR% estimates of cancer have shown large variation across populations, methods, and data sources.

Methods

We evaluated variations in PAR% of postmenopausal breast cancer by obesity, alcohol, fruit and vegetable intake, and physical activity in the Nurses' Health Study (NHS). For the low-risk method, the age-standardized incidence rate in the low-risk group was compared with the overall rate. The partial PAR% method estimated the PAR% by the risk factors while adjusting for other covariates. The conventional method estimated the PAR% based on the exposure distribution and relative risks in the NHS, or from US national surveys and meta-analyses. For each model, we compared the results obtained using baseline or repeated measurements.

Results

When considering the risk factors simultaneously, within the NHS, the low-risk method yielded overall PAR% for the baseline, simple update, and cumulative average models of 17.4%, 25.2%, 29.3%; the partial PAR% method yielded estimates at 13.7%, 28.0%, 31.2%. When considering each risk factor individually and then combined, the estimates by the partial PAR% were 12.5%, 20.7%, 18.9%; while the conventional method gave corresponding PAR% of 13.8%, 21.1%, 18.6%. The estimated PAR% based on meta-analyses and US national surveys was 25.6%.

Conclusions

The three methods provided similar PAR% based on the same data source, timing of measurements, and target populations. However, sizable increases in the PAR% were observed for repeated measures over a single measure and for calculations based on achieving all recommendations jointly rather than individually.

Introduction

Cancer is a leading cause of disease burden and mortality across countries [1, 2], while a substantial proportion could be prevented by primary intervention[3]. The population attributable risk (PAR%) estimates the percentage of disease in a target population that theoretically would not have occurred through the optimization of its etiologic factors.

There are various methods to calculate PAR%. The low-risk method compares the age-standardized incidence rate for the disease for the low-risk group to the rate in the entire study population[3]. The partial PAR% method accounts for the joint distribution of modifiable and non-modifiable risk factors and estimates an attainable preventable fraction [4]. The conventional method incorporates exposure prevalence from national or regional representative populations, and relative risks from published literature. It does not normally consider multiple factors jointly, but rather combine the individual estimates under the independent assumption of the risk factors[5]. Yet, it is unknown if the effect would be stronger if the population adhered to all recommendations.

For each method, the data sources can also differ. To estimate the prevalence, some studies have used cross-sectional surveys [6–8], while others used empirical data from cohort studies [4]. To estimate the relative risks, some extracted data from published literature, some from observational studies with one exposure assessment [7], while others used repeated measurements [4]. Results from studies based on a single measure are more prone to measurement error; if the relative risk is underestimated, the PAR% likewise will be underestimated.

We evaluated in the Nurses' Health Study (NHS) the degree to which PAR% of postmenopausal breast cancer 1. by the statistical methods; 2. by using single versus repeated measurements; 3. by considering exposures individually or in combination; and 4. by the sources of the relative risk and prevalence data. We calculated the PAR% of four lifestyle risk factors in the 2018 World Cancer Research Fund and American Institute for Cancer Research (WCRF/AICR) cancer prevention recommendations for which the evidence for an association with breast cancer was strong (alcohol consumption, body mass index [BMI], and physical activity) or suggestive (fruit and vegetable intake) [9]. Dairy and calcium intake was considered as suggestive risk factors by WCRF/AICR, but were not included based on the null result from a recent comprehensive assessment in the largest international consortium of diet and breast cancer [10].

Methods

Nurses' Health Study

The NHS was established in 1976 when 121,701 female nurses aged 30–55 years returned the initial questionnaire [11]. Participants have been followed up biennially to collect their medical, lifestyle, and other health-related information. The overall response rate has been greater than 90%.

Assessment of exposures

In 1980, 1984, 1986 and every 4 years thereafter, participants returned semiquantitative food frequency questionnaire (FFQ) covering their usual diet in the past year. Fruit and vegetable consumption was estimated based on the quantity and frequency of all relevant food items consumed and, for alcohol, on the alcohol content of the beverages consumed [12,13]. Studies have documented moderate to high validity (**Supplementary methods**) of the questionnaire in measuring intake of alcohol and fruit and

vegetables [14,15]. Starting in 1986 and every 2–4 years thereafter, participants reported their average time per week spent engaging in various types of physical activity [16]. Total physical activity was converted into metabolic equivalent task hours per week [17]. Information on height was collected in 1976. Weight was reported in 1976 and biennially afterwards [18].

Assessment of covariates

Age at menarche was collected in 1976. Menopausal status and history of benign breast disease were assessed at baseline and updated biennially. Family history of breast cancer was obtained in 1982 and updated every 4 years beginning in 1988. Oral contraceptive use was assessed in 1980, 1982, and 1984. Age at first birth was asked updated biennially until 1982. Parity was asked biennially until 1996. Postmenopausal hormone therapy (HRT) use was asked biennially until 2004.

Case ascertainment in NHS

Participants in the NHS were asked to report the diagnosis of invasive breast cancer biennially. Medical records and pathology reports were retrieved upon participant permission (or next of kin for those who had died) to confirm the identified diagnosis. Only postmenopausal breast cancer cases were included in the analysis.

US national exposure prevalence data

US national distributions for alcohol consumption were obtained from the National Health Interview Survey (NHIS, 2013 and 2014)[19]. Information on BMI, fruit and vegetable intake and physical activity were obtained from the National Health and Nutrition Examination Survey (NHANES, 2011–2012, 2013–2014) [20].

Risk estimates for lifestyle factor and breast cancer associations from meta-analyses

Relative risk (RR) estimates for each risk factor were identified from the most updated meta-analysis on their association with risk of breast cancer incidence. When the cut points in the meta-analysis did not match the national surveys, we imposed a log-linear relationship on the RR from continuous analysis and calculated the RR for each level of the exposure.

Statistical analysis

1986, when physical activity was first asked in the NHS, was considered as the baseline. Participants were excluded if they had a history of cancer (except for nonmelanoma skin cancer), missing value of the main exposures at baseline, or had extreme total energy intake (below 600 or above 3,500 kcal/day).

We calculated RRs of postmenopausal breast cancer in which each exposure was modeled as binary variables representing high- (listed first) and low-risk categories based on the WCRF/AICR cancer prevention recommendations[21]: alcohol consumption (drinker vs non-drinker), BMI (≥ 25 vs < 25 kg/m^2), fruit and vegetable intake (< 5 vs ≥ 5 servings/day), and physical activity (< 18 vs ≥ 18 MET-

hours/week). Value from the last questionnaire returned was carried forward where missing. Age and multivariable-adjusted relative risks (and 95% confidence intervals [CI]) were estimated using Cox proportional hazards models. For models using repeated measurements, exposures up to the assessment just before diagnosis, loss to follow-up, or the last assessment before the end of follow-up were used, and the covariates were included as time-varying variables whenever possible. Specific categorizations of the covariates included in the multivariable models can be found in **Table 4**.

To compare PAR% computed by the statistical methods and by sources of relative risk and prevalence data, we designed 6 model sets, each including three models for the different timings in the NHS (**Supplementary methods**). A model using the conventional method, published meta-analyses for the relative risks, and the US national survey prevalence data was also constructed to compare with the results from NHS. Figure 1 summarizes the methods and comparisons used in each model.

Low-risk method

The low-risk group was defined as those being at the optimal level for all four risk factors. Age-specific postmenopausal breast cancer incidence rates in NHS were calculated for the low-risk group and the entire cohort for 5-year age groups, then standardized by the number of people in each age group. PAR% (Model 1) was calculated as:

PAR% = (age-standardized incidence rate in NHS - rate in the low-risk group) / rate in NHS

We calculated the PAR% in the US population (Model 2) as:

PAR% = (US age-standardized incidence rate over 55-year-old* - rate in the low-risk group) / US age-standardized incidence rate over 55-year-old

*at which age over 90% of US women would have become postmenopausal [22,23].

In sensitivity analyses, we limited the study population to non-users of HRT, including never users and those who took HRT for less than 10 years and stopped using it for four years or longer[24].

Partial PAR% method

The partial PAR% reflects the proportional reduction expected in disease incidence if all the risk factors of interest were set to the optimal level while holding the other covariates unchanged. We calculated partial PAR% and 95% CIs using the %PAR SAS macro by Spiegelman et al[25]. Briefly, it splits the study population into all possible combinations of the exposures and covariate categories and calculated the excessive risk for each stratum compared to the low-risk group (**Supplementary methods**).

We calculated the partial PAR% using two approaches. One (Model 3) classified people into the high- and low-risk groups using the same criteria described in the low-risk method and obtained the PAR% of the overall high-risk affiliation. The other (Model 4) took in the four risk factors into the model to obtain a composite partial PAR% for being at elevated risk for each of the factors [25].

In sensitivity analyses, we substituted BMI with adult weight change to examine the PAR%. We also corrected for fruit and vegetable intake relative to dietary records[26] to evaluate the bias due to over-reporting.

Conventional method

For each risk factor, we applied the modified Levin's formula[27] to the effect size (RR_n) and the prevalence (P_n) for the *n-th* category (Model 5–7):

$$\mathrm{IndividualPAR\%} = \frac{\sum_{n=1}^{N} P_n * (RR_n - 1)}{\sum_{n=1}^{N} P_n * (RR_n - 1) + 1}.$$

The overall PAR% was then computed by applying the multiplicative formula, under the assumption of independent exposures and effects under study[5]:

$$\mathrm{OverallPAR\%} = 1 - \prod_{k=1}^{K} (1 - PAR_k\%)$$

where PAR_k% denotes the PAR% of each risk factor *k*.

For all hypothesis tests, a *p-value*< 0.05 was considered statistically significant, and all tests of statistical significance were 2-sided. Analyses of the conventional method were conducted using R, version 3.6.0 (R Foundation for Statistical Computing, Vienna, Austria.). All other analyses were conducted using SAS, version 9.4 (SAS Institute, Inc., Cary, NC).

Results

Compared with the high-risk group, women in the low-risk group consumed no alcohol, had lower BMI, higher fruit and vegetable intake, and higher physical activity at baseline. The mean age, height, and age at menarche, distribution of parity and age at first birth, the proportion with a personal history of benign breast disease or family history of breast cancer, and menopausal status/menopausal HRT use were similar between the groups. Women in the low-risk group were less likely to be current smokers (Table 1). Fewer women were classified as non-drinkers based on cumulative average intake over follow-up as compared to baseline. Throughout study follow-up, fruit and vegetable intake remained about the same, more women had higher BMI, and total physical activity tended to increase (**Table S1**).

We identified 6708 incident invasive postmenopausal breast cancers during 28 years of follow-up in the NHS. Higher alcohol consumption, higher BMI, lower fruit and vegetable intake were associated with higher risk of postmenopausal breast cancer. Low physical activity was marginally associated with higher risk of postmenopausal breast cancer (**Table S2**).

PAR% estimated by the low-risk method

Table 3 shows the age-standardized incidence rate for postmenopausal breast cancer in the low-risk group in NHS, in all NHS participants, and in the general US population. The rate for the low-risk group ranged from 267.9 to 312.7 per 100,000 person-years depending on whether data collected at baseline or updated during follow-up were used. The rate for the entire NHS cohort was slightly higher (378.8/100,000 person-years, 1986–2014) than that for the US general population (358.2/100,000 person-years, 1975–2017), likely resulting from differences in other risk factors (e.g., HRT use, reproductive variables) or breast cancer screening.

The PAR% for the low-risk group in NHS was 17.4% when using only baseline data, 25.2% for the simple updated exposures, and 29.3% for the cumulative averaged exposures. The PAR% for the low-risk group defined using US population prevalence were 12.7%, 20.8%, and 25.2%, respectively. In sensitivity analyses restricting the population to non-HRT users, the estimated PAR% remained similar (data not shown).

PAR% estimated by the partial PAR% method

Differences in PAR% were observed when the exposures at baseline, simple update, or cumulative average were used. Compared to the models of repeated measures, the strengths of the associations tended to be weaker when baseline information was used (**Table 4**). The proportion of total person-time allocated to high-risk categories was generally higher in models with repeated measurements, except for physical activity.

Taken together, had everyone switched their alcohol consumption, BMI, fruit and vegetable intake, and physical activity to the low-risk levels, the estimated preventable fraction was 12.5% if baseline information was used, 20.7% for the simple update model, and 18.9% for the cumulative average model, holding the other covariates unchanged (**Table 4**). The corresponding PAR% estimated by the overall risk classification were 13.7%, 28.0% and 31.2%. In sensitivity analysis where we substituted BMI with adult weight change, the overall corresponding estimated PAR%'s were even higher: 22.5%, 35.5%, and 38.3% (data not shown).

PAR% by the conventional method

The overall PAR% of postmenopausal breast cancer calculated using the conventional method was computed using RRs and prevalence data from NHS (**Table S2**) or from NHANES and NHIS (Table 2), and summary RRs from meta-analyses (**Table S4)**. When using RRs and prevalence from the NHS, the estimated PAR% was 13.8% for the baseline model, 21.1% for the simple update model, and 18.6% for the cumulative average model (Fig. 3). When using RRs from the NHS and prevalence data from NHANES and NHIS, the PAR%'s followed the same pattern but were higher by ~ 7% due to the higher prevalence of the high-risk group in the US. Using RRs from meta-analyses and prevalence data from NHANES and NHIS, which resembled PAR% calculations based on published literature, the estimated PAR% was 25.6%, falling between the estimated PAR% by the two models above.

Discussion

Currently published PAR%'s of breast cancer by a combination of modifiable risk factors range from 26.0–40.7% [4,6,8,28–31] (**Table S5**). Contributing to the variation is the varying distribution of the risk factors under study; for example, a previous analysis in the NHS used repeated measures of alcohol, weight change since age 18 years, physical activity, breastfeeding, and HRT use, yielding a PAR% of 34.6%[4]. Besides different factors considered, the data sources, modeling of the risk factors, and statistical methods could also contribute to the variation. Further, many PAR% estimates in the literature computed the full PAR% which combined modifiable and non-modifiable risk factors, some of which are unrealistic to intervene upon. To investigate the sources of variation, we estimated the preventability of postmenopausal breast cancer associated with four modifiable risk factors using three statistical methods, with three timings of exposure measurements, and considering whether low risk groups were defined for risk factors individually or combined.

Little influence by choice of statistical method

For the overall PAR% for these four risk factors, the preventable fraction did not differ much by the choice of statistical method given the same timing of exposure assessment. The low-risk method and the partial PAR% method generated similar estimates (Fig. 2, top 2 panels). This reassured the robustness of the methods when the non-modifiable covariates were adjusted inherently in the low-risk method and analytically in the partial PAR% method. When estimating the preventability by individual risk factors and then combining them (i.e., the conventional method), the results were comparable (Fig. 2, bottom 2 panels). Although the correlations between the risk factors might lead to over- or underestimated PAR%, the similarity of the results suggests little bias due to correlated healthy behaviors. Thus, the accuracy of the data sources is more critical than the statistical method used.

Repeated measures demonstrated higher PAR% than a single measurement

Within each model set, the overall PAR% were generally higher by at least 50% when repeated measures were used, mainly because the repeated measurements capture the level of time-varying exposures more precisely, therefore the RRs were greater in magnitude and the corresponding prevalence of the low-risk group were lower (Fig. 2, all panels). Moreover, repeated measures of lifestyle factors are time-integrated, indicating the preventability of sustained behavioral changes. In summary, the same pattern seen in results by every method suggests that PAR% might be underestimated by as much as 50% in cohort studies with single baseline measures.

Combining healthy lifestyle factors yields greater preventability estimates

Methods defining the low-risk group by combining risk factors yielded substantially greater PAR%s than methods that considered the low-risk group for each risk factor individually and then combined the results (Fig. 2, top 2 versus bottom 2 panels). As discussed above, the comparable results from the low-risk and conventional methods suggest little influence by correlation between the risk factors. The difference in magnitude is more likely that the combined low-risk group has more extreme healthy behaviors. For example, in the individual method, the low-risk group for BMI would include all individuals with BMI < 25 kg/m^2. In the combined low risk group, the individuals must have BMI < 25 kg/m^2 and also more activity, less alcohol and more fruits and vegetables. In this study, the low-risk population defined by the risk factors jointly on average consumed 0.5 more servings of fruits and vegetables per day, engaged in 3 more MET-hours of activity per week, and attained 50% less weight gain in adulthood compared to when each factor was defined individually. Although it is difficult to achieve multiple healthy behaviors, it has been shown that success in achieving one in an overall healthy lifestyle facilitated improvements in others. [32] The WCRF/AICR recommendations are intended to be a package, so the PAR% using the combined low-risk group is within the spirit of what can be theoretically achieved by adhering to all.

There are limitations in our study. Measurement error was a concern - although the systematic over-reporting did not affect the relative ranking of the participants, it did change the prevalence of each category. In sensitivity analysis, we corrected for the over-reporting in fruit and vegetable intake as an example to evaluate how much it could affect the PAR%[26]. After calibration, when using the conventional methods, the estimated PAR% in NHS went up by 4% but remained lower than that estimated for the general US population. Lastly, the study population consists of predominately White female nurses, which may limit the validity of extrapolation to other populations. Nevertheless, the main purpose of this study is not to quantify the preventability, but to demonstrate the sources and degree of variation in the computation.

Our results suggest that the three statistical methods, given the same data source, provide similar results. However, sizable increases in the PAR% were observed when repeated measures were used instead of a single measure, as well as for calculations based on women achieving all recommendations rather than considering each individually and then combining the results. Thus, PAR%s in the current literature have likely underestimated the preventable fraction of postmenopausal breast cancer. These results emphasize the importance of high-quality data sources, call for cautious interpretation of PAR% in the current literature, and underscore the potential of long-term behavioral change toward an overall healthier lifestyle to optimally lower postmenopausal breast cancer risk.

Declarations

Funding

This work was supported by World Cancer Research Fund UK (WCRF) [grant 2018/1818 to E.L. Giovannucci] as part of the World Cancer Research Fund International grant programme; and by the U.S. National Institutes of Health [UM1 CA186107, U01 CA176726, P01 CA87969, and R01 CA50385].

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author contributions

Edward Giovannucci and You Wu contributed to the study conception and design. The study methodology was developed by Edward Giovannucci, You Wu, Kim Hanseul, Kai Wang, and Molin Wang. The formal analysis was conducted by You Wu. Edward Giovannucci, Walter Willett, and Stephanie Smith-Warner supervised the study. The first draft of the manuscript was written by You Wu and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethics Approval

The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Acknowledgement

The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) and/or the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. Central registries may also be supported by state agencies, universities, and cancer centers. Participating central cancer registries include the following: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Indiana, Iowa, Kentucky, Louisiana, Massachusetts, Maine, Maryland, Michigan, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico, Rhode Island, Seattle SEER Registry, South Carolina, Tennessee, Texas, Utah, Virginia, West Virginia, Wyoming.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424.
2. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392:1736–88.

3. Song M, Giovannucci E. Preventable Incidence and Mortality of Carcinoma Associated With Lifestyle Factors Among White Adults in the United States. JAMA Oncol. 2016;2:1154–61.
4. Tamimi RM, Spiegelman D, Smith-Warner SA, Wang M, Pazaris M, Willett WC, et al. Population Attributable Risk of Modifiable and Nonmodifiable Breast Cancer Risk Factors in Postmenopausal Breast Cancer. American Journal of Epidemiology. 2016;184:884–93.
5. Ezzati M, Vander Hoorn S, Rodgers A, Lopez AD, Mathers CD, Murray CJ, et al. Estimates of global and regional potential health gains from reducing multiple major risk factors. Lancet. 2003;362:271–80.
6. Wilson LF, Page AN, Dunn NAM, Pandeya N, Protani MM, Taylor RJ. Population attributable risk of modifiable risk factors associated with invasive breast cancer in women aged 45-69 years in Queensland, Australia. Maturitas. 2013;76:370–6.
7. Islami F, Goding Sauer A, Miller KD, Siegel RL, Fedewa SA, Jacobs EJ, et al. Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. CA Cancer J Clin. 15 ed. 2018;68:31–54.
8. Hayes J, Richardson A, Frampton C. Population attributable risks for modifiable lifestyle factors and breast cancer in New Zealand women. Intern Med J. 2013;43:1198–204.
9. World Cancer Research Fund/American Institute for Cancer Research. World Cancer Research FundAmerican Institute for Cancer Research. Continuous Update Project Expert Report 2018.
10. Wu Y, Huang R, Wang M, Bernstein L, Bethea TN, Chen C, et al. Dairy foods, calcium, and risk of breast cancer overall and for subtypes defined by estrogen receptor status: a pooled analysis of 21 cohort studies. Am. J. Clin. Nutr. 2021;114:450–61.
11. Colditz GA, Martin P, Stampfer MJ, Willett WC, Sampson L, Rosner B, et al. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. American Journal of Epidemiology. 1986;123:894–900.
12. Hirko KA, Chen WY, Willett WC, Rosner BA, Hankinson SE, Beck AH, et al. Alcohol consumption and risk of breast cancer by molecular subtype: Prospective analysis of the nurses' health study after 26 years of follow-up. Int J Cancer. 2016;138:1094–101.
13. Farvid MS, Chen WY, Rosner BA, Tamimi RM, Willett WC, Eliassen AH. Fruit and vegetable consumption and breast cancer incidence: Repeated measures over 30 years of follow-up. Int J Cancer. 2019;144:1496–510.
14. Salvini S, Hunter DJ, Sampson L, Stampfer MJ, Colditz GA, Rosner B, et al. Food-based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. Int J Epidemiol. 1989;18:858–67.
15. Giovannucci E, Colditz G, Stampfer MJ, Rimm EB, Litin L, Sampson L, et al. The assessment of alcohol consumption by a simple self-administered questionnaire. American Journal of Epidemiology. 1991;133:810–7.
16. Al-Shaar L, Pernar CH, Chomistek AK, Rimm EB, Rood J, Stampfer MJ, et al. Reproducibility, Validity, and Relative Validity of Self-Report Methods for Assessing Physical Activity in Epidemiologic

Studies: Findings from the Women's Lifestyle Validation Study. American Journal of Epidemiology. 2022.

17. Eliassen AH, Hankinson SE, Rosner B, Holmes MD, Willett WC. Physical Activity and Risk of Breast Cancer Among Postmenopausal Women. Arch. Intern. Med. 2010;170.
18. Rimm EB, Stampfer MJ, Colditz GA, Chute CG, Litin LB, Willett WC. Validity of self-reported waist and hip circumferences in men and women. Epidemiology. 1990;1:466–73.
19. Breslow RA, Castle I-JP, Chen CM, Graubard BI. Trends in Alcohol Consumption Among Older Americans: National Health Interview Surveys, 1997 to 2014. Alcohol Clin Exp Res. 2017;41:976–86.
20. Dwyer J, Picciano MF, Raiten DJ, Members of the Steering Committee, National Health and Nutrition Examination Survery. Collection of food and dietary supplement intake data: What We Eat in America-NHANES. Journal of Nutrition. 2003;133:590S–600S.
21. World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report 2018. Recommendations and public health and policy implications.
22. MacMahon B, Worcester J. Age at menopause. United States--1960-1962. Vital Health Stat 11. Vital Health Stat 11; 1966;:1–20.
23. Nichols HB, Trentham-Dietz A, Hampton JM, Titus-Ernstoff L, Egan KM, Willett WC, et al. From menarche to menopause: trends among US Women born from 1912 to 1969. American Journal of Epidemiology. 2006;164:1003–11.
24. Collaborative Group on Hormonal Factors in Breast Cancer. Type and timing of menopausal hormone therapy and breast cancer risk: individual participant meta-analysis of the worldwide epidemiological evidence. Lancet. 2019;394:1159–68.
25. Spiegelman D, Hertzmark E, Wand HC. Point and interval estimates of partial population attributable risks in cohort studies: examples and software. Cancer Causes Control. Kluwer Academic Publishers; 2007;18:571–9.
26. Feskanich D, Rimm EB, Giovannucci EL, Colditz GA, Stampfer MJ, Litin LB, et al. Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. J Am Diet Assoc. 1993;93:790–6.
27. Hanley J. A heuristic approach to the formulas for population attributable fraction. J Epidemiol Community Health. 2001;55:508–14.
28. Sprague BL, Trentham-Dietz A, Egan KM, Titus-Ernstoff L, Hampton JM, Newcomb PA. Proportion of invasive breast cancer attributable to risk factors modifiable after menopause. American Journal of Epidemiology. 2008;168:404–11.
29. Barnes BBE, Steindorf K, Hein R, Flesch-Janys D, Chang-Claude J. Population attributable risk of invasive postmenopausal breast cancer and breast cancer subtypes for modifiable and non-modifiable risk factors. Cancer Epidemiol. 2011;35:345–52.
30. van Gemert WA, Lanting CI, Goldbohm RA, van den Brandt PA, Grooters HG, Kampman E, et al. The proportion of postmenopausal breast cancer cases in the Netherlands attributable to lifestyle-related risk factors. Breast Cancer Res. Treat. 3rd ed. 2015;152:155–62.

31. Clarke CA, Purdie DM, Glaser SL. Population attributable risk of breast cancer in white women associated with immediately modifiable risk factors. BMC CANCER. 2006;6:170.
32. Lippke S, Nigg CR, Maddock JE. Health-promoting and health-risk behaviors: theory-driven analyses of multiple health behavior change in three international samples. Int J Behav Med. 2012;19:1–13.

Tables

Table 1. Baseline characteristics of participants according to high and low risk status of postmenopausal breast cancer[a] in the Nurses' Health Study

	Low-risk group	High-risk group
	Mean (SD) or percentage (%)	
Risk factors		
Alcohol consumption, grams/d	0 (0)	6.3 (10.8)
Body mass index (BMI), kg/m^2	22.1 (1.8)	25.4 (4.8)
Fruit and vegetable intake, servings/d	7.7 (2.7)	4.9 (2.6)
Physical activity, MET-hrs[b]/week	40.3 (28.1)	13.5 (20.2)
Covariates		
Age, years	53.2 (7.4)	52.5 (7.2)
Height, m	1.64 (0.06)	1.64 (0.06)
Age at menarche, years	12.6 (1.8)	12.4 (1.8)
Race		
White	95.9%	97.8%
Parity and age at first birth (years)		
Nulliparous	6.5%	5.8%
1-2 children age at 1st birth <25	14.8%	13.9%
1-2 children age at 1st birth 25-29	17.0%	15.3%
1-2 children age at 1st birth ≥30	6.9%	6.2%
3-4 children age at 1st birth <25	23.1%	25.8%
3-4 children age at 1st birth 25-29	17.2%	16.3%
3-4 children age at 1st birth ≥30	2.4%	2.6%
≥5 children	7.1%	9.2%
Missing parity or age at 1st birth	5.0%	4.8%
Oral contraceptive (OC) use		
No OC use	59.2%	52.9%
>0-2 yrs of OC use	15.7%	17.9%
>2-5yrs of OC use	11.3%	12.9%
>5-10 yrs of OC use	10.1%	11.2%

>10 yrs of OC use	3.7%	5.0%
History of benign breast disease	31.2%	29.8%
Family history of breast cancer	8.0%	8.1%
Menopausal hormone therapy use		
Premenopausal/unknown menopausal status	30.7%	33.4%
Never user among postmenopausal women	33.8%	35.4%
Current user among postmenopausal women	19.6%	17.0%
Past user among postmenopausal women	15.8%	14.3%
Smoking status		
Never smoker	58.9%	43.8%
Past smoker	28.2%	34.9%
Current smoker	13.0%	21.3%
[a] Low risk status for breast cancer defined by meeting the following four criteria: alcohol consumption (0g/day), body mass index (<25kg/m^2), fruit and vegetable intake (≥5 servings/day), and physical activity (≥18 total MET-hrs/week). Others were defined as having high risk.		
[b] MET-hrs, metabolic equivalent of task per hour, a measure of relative intensity of different physical activities as compared to the resting metabolic rate.		

Table 2. Distribution[a] of the risk factors[b] for postmenopausal breast cancer in the Nurses' Health Study (NHS), the National Health and Nutrition Examination Survey (NHANES), and the National Health Interview Survey (NHIS)

	NHS				NHANES & NHIS[c]	
	Category	Prevalence (%)			Category[f]	Prevalence (%)
		Baseline	Simple update[d]	Cumulative average[e]		
Alcohol consumption						
	0 gram/day	35.1	42.5	27.2	**Nondrinker**	40.7
	>0.0-13.9 grams/day	51.2	44.5	60.5	< 1 drink/day	30.7
	14.0-27.9 grams/day	8.2	8.2	8.4	1-<2 drinks/day	8.1
	≥28 grams/day	5.5	4.8	3.9	≥2 drinks/day	13.6
Body mass index						
	< 25 kg/m^2	57.4	47.1	53.4	**< 25 kg/m^2**	28.7
	25-29.9 kg/m^2	28.5	32.6	31.7	25-29.9 kg/m^2	29.6
	≥30 kg/m^2	14.1	20.3	15.0	≥30 kg/m^2	40.1
Fruit and vegetable intake[g]						
	0-<2 servings/day	8.5	9.9	5.8	0-199 grams/day	49.2
	2-<5 servings/day	51.3	52.3	54.5	200-399 grams/day	41.3
	≥5 servings/day	40.3	37.8	39.7	**≥400 grams/day**	9.4
Physical activity						
	<4 MET-hrs[h]/week	34.1	27.9	17.3	<4.2 MET-hrs/week	57.9
	4-<12.5 MET-hrs/week	30.2	27.3	33.9	4.2-<12.5 MET-hrs/week	15.6
	12.5-<18 MET-hrs/week	9.6	11.4	15.2	12.5-<16.7 MET-hrs/week	5.4
	≥18 MET-hrs/week	26.0	33.4	33.6	**≥16.7 MET-hrs/week**	21.1

[a] Theoretical minimum risk (reference level) in bold

[b] Risk factors were selected based on the 2018 World Cancer Research Fund and American Institute for Cancer Research (WCRF/AICR) cancer prevention recommendations[21] for which the evidence for an association with breast cancer was considered convincing, probable, or suggestive

[c] US national exposure distributions for alcohol consumption were obtained from the National Health Interview Survey (NHIS 2013 and 2014), while information on BMI, fruit and vegetable intake and physical activity were obtained from the National Health and Nutrition Examination Survey (NHANES 2011-2012, 2013-2014).

[d] The measurement from the most recent questionnaire prior to breast cancer diagnosis, death, lost to follow-up or administrative end of follow-up for each questionnaire return cycle.

[e] The average of all past measurements for each follow-up cycle.

[f] The equivalent cut points for low- and high-risk groups, respectively, from the US prevalence data were: alcohol consumption (>0 vs 0 gram/day), BMI (≥25 vs <25 kg/m^2), fruit and vegetable intake (<400 vs ≥400 grams/day), physical activity (<1000 vs ≥1000 MET-min/week).

[g] Sample average weight of fruits and vegetables: one apple (80g), 6 strawberries (100g), one orange (120g); 10 medium lettuce leaves (75g), 5 broccoli florets (100g), one medium carrot (60g).

[h] MET, metabolic equivalent of task, a measure of relative intensity of different physical activities as compared to the resting metabolic rate.

Table 3. Age-standardized incidence rates of postmenopausal breast cancer in the low-risk group [a] and the overall Nurses' Health Study (NHS) population, and the US population, and the corresponding estimated population attributable risk (PAR%) according to the timing of the exposure measurements

Timing of the exposure measurements [b]	Incidence rate in the low-risk group in NHS (per 100,000 person-year)	Incidence rate in NHS (per 100,000 person-year)	Incidence rate in the US population (per 100,000 person-year)	PAR% (NHS)	PAR% (US population)
Baseline	312.7	378.8	358.2	17.4%	12.7%
Simple update	283.5	378.8		25.2%	20.8%
Cumulative average	267.9	378.8		29.3%	25.2%

[a] Low risk status defined by meeting all four criteria of alcohol consumption (0g/day), body mass index (<25kg/m2), fruit and vegetable intake (≥5 servings/day), and physical activity (≥18 total MET-hr/week).

[b] (1) Baseline: using measurements of the exposures at study enrollment in 1986 (or 1990 for women who did not complete the 1986 FFQ); (2) Simple update: using the measurement of the exposures from the assessment just prior to diagnosis, death, lost to follow-up or administrative end; and (3) Cumulative average: using average exposure calculated from the assessment just prior to diagnosis, death, lost to follow-up or administrative end and all prior questionnaire cycles.

Table 4. Prevalence, multivariable[a] relative risk (RR) and 95% confidence interval (CI) for postmenopausal breast cancer, and partial population attributable risk (PAR%) for alcohol consumption, body mass index (BMI), fruit and vegetable intake, and physical activity in the Nurses' Health Study

		Baseline	Simple update [b]	Cumulative average[c]
High and low risk groups were defined for each risk factor individually				
	No. of cases	6708	6708	6708
Alcohol consumption (>0g/day)				
	High-risk%	63.6%	57.4%	72.9%
	RR[5] (95% CI)	1.07 (1.02-1.13)	1.12 (1.06-1.18)	1.11 (1.05-1.17)
BMI (≥25kg/m^2)				
	High-risk%	42.6%	52.9%	52.9%
	RR (95% CI)	1.15 (1.09-1.21)	1.19 (1.13-1.25)	1.19 (1.13-1.25)
Fruit and vegetable intake (<5 servings/day)				
	High-risk%	59.7%	62.2%	60.3%
	RR (95% CI)	1.04 (0.99-1.09)	1.08 (1.03-1.14)	1.05 (1.00-1.11)
Physical activity (<18 total MET-hr/week)				
	High-risk%	74.0%	66.6%	66.4%
	RR (95% CI)	1.01 (0.96-1.07)	1.04 (0.98-1.09)	1.02 (0.97-1.07)
Total	Overall PAR%	12.5%	20.7%	18.9%
		(1.6% - 23.2%)	(10.7% - 30.3%)	(7.9% - 29.5%)
High and low risk groups were defined by the joint classification of the four risk factors				
	No. of cases	6708	6708	6708
High-risk group[d]	Prevalence	97.8%	97.2%	98.1%
	RR (95% CI)	1.16 (0.98-1.38)	1.40 (1.18-1.65)	1.46 (1.19-1.80)
	PAR%	13.7%	28.0%	31.2%

	(-1.2% - 28.1%)	(15.8% - 39.3%)	(16.5% - 44.5%)	
[a] Multivariable model includes: age (<50, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, ≥80 years), height (<1.60, 1.60-1.64, 1.65-1.69, 1.70-1.74, ≥1.75m), age at menarche (<12, 12, 13, 14, >14 years), duration of oral contraceptive use (no use, >0-2, >2-5, >5-10, >10 years), joint classification of age at first birth (AFB) and parity (nulliparous, 1-2 children and AFB <25, 1-2 children and AFB 25-29, 1-2 children and AFB ≥30, 3-4 children and AFB <25, 3-4 children and AFB 25-29, 3-4 children and AFB 30+, ≥5 children and AFB <25, ≥5 children and AFB ≥25), menopausal status and menopausal hormone therapy use (premenopausal/unknown menopausal status, never user among postmenopausal women, current user among postmenopausal women, past user among postmenopausal women), history of benign breast disease (yes, no), family history of breast cancer (yes, no), total energy intake (kcal/d, quintiles). Missing value for the covariates were filled in by carrying-forward responses from the last questionnaire for analyses using repeated measurements.			
[b] The measurement from the most recent questionnaire return for each follow-up cycle.				
[c] The average of all past measurements for each follow-up cycle.				
[d] Low-risk is defined as being at the optimal level for all four risk factors for alcohol, BMI, fruit and vegetable, and physical activity while all other participants were considered to be at high-risk.				
[e] RR, relative risk.				

Figures

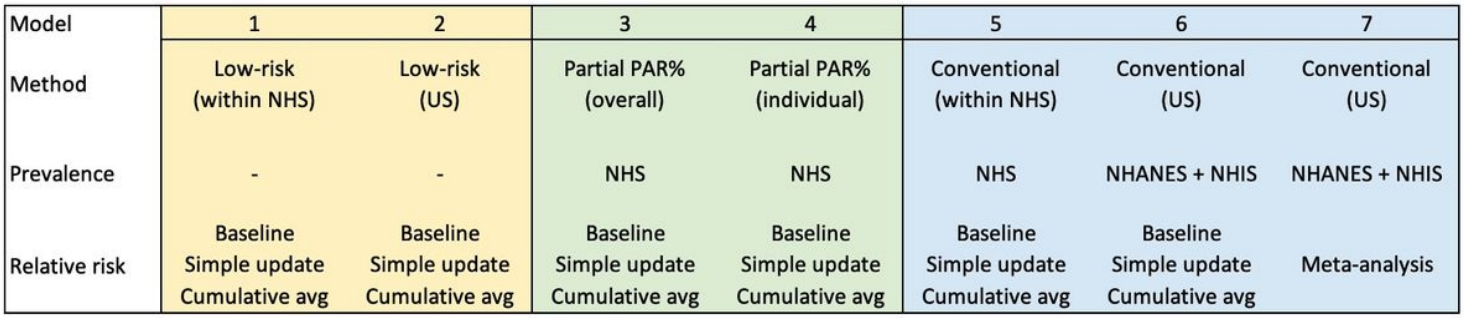

Figure 1

Summary of the methods, sources of relative risks, prevalence, and timing of measurements

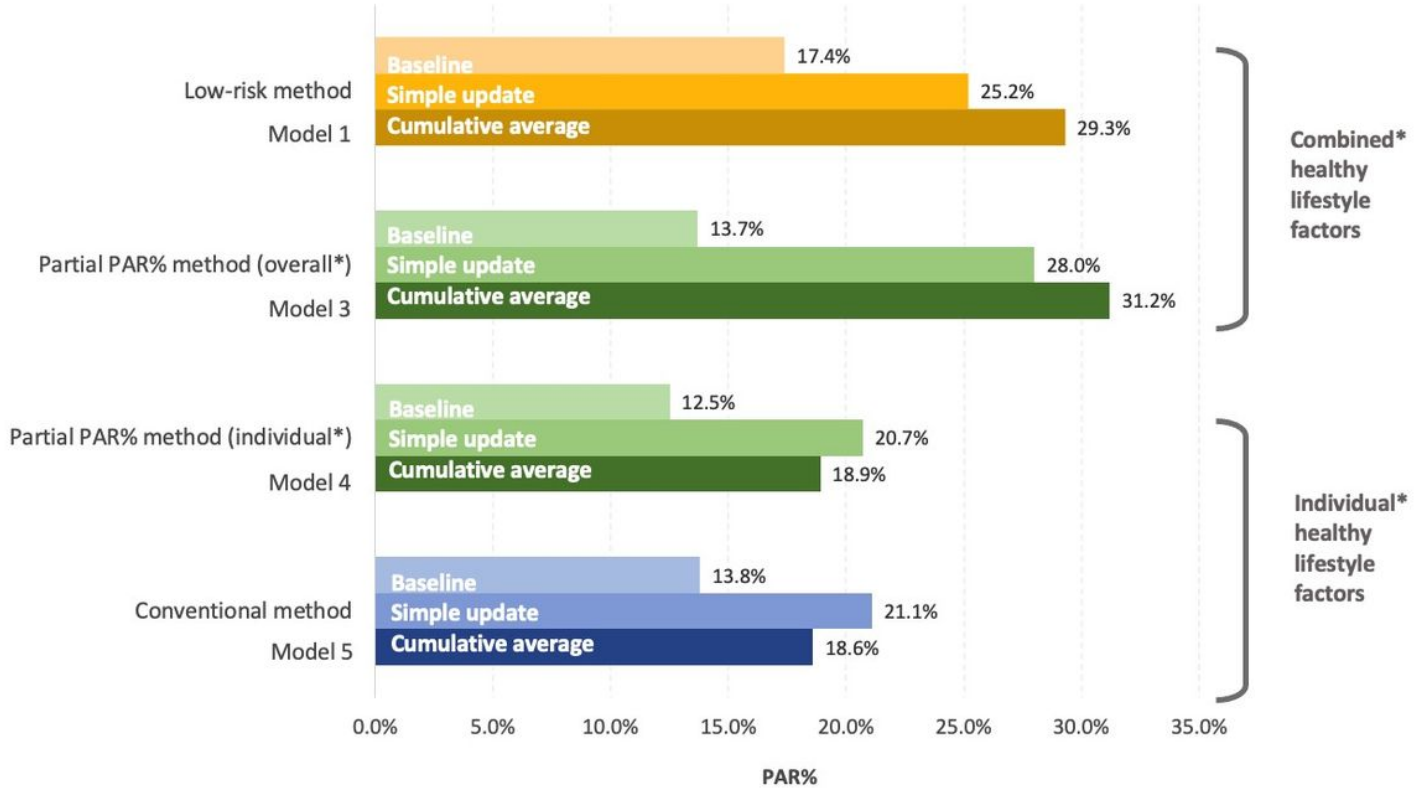

Figure 2

PAR% of postmenopausal breast cancer in the Nurses' Health Study by alcohol, body mass index, and fruit and vegetable intake, and physical activity. Combined: considering the four risk factors simultaneously for low-risk classification to obtain an overall PAR%; individual: considering the four risk factors separately and then combined the individual PAR% to obtain a composite partial PAR%.

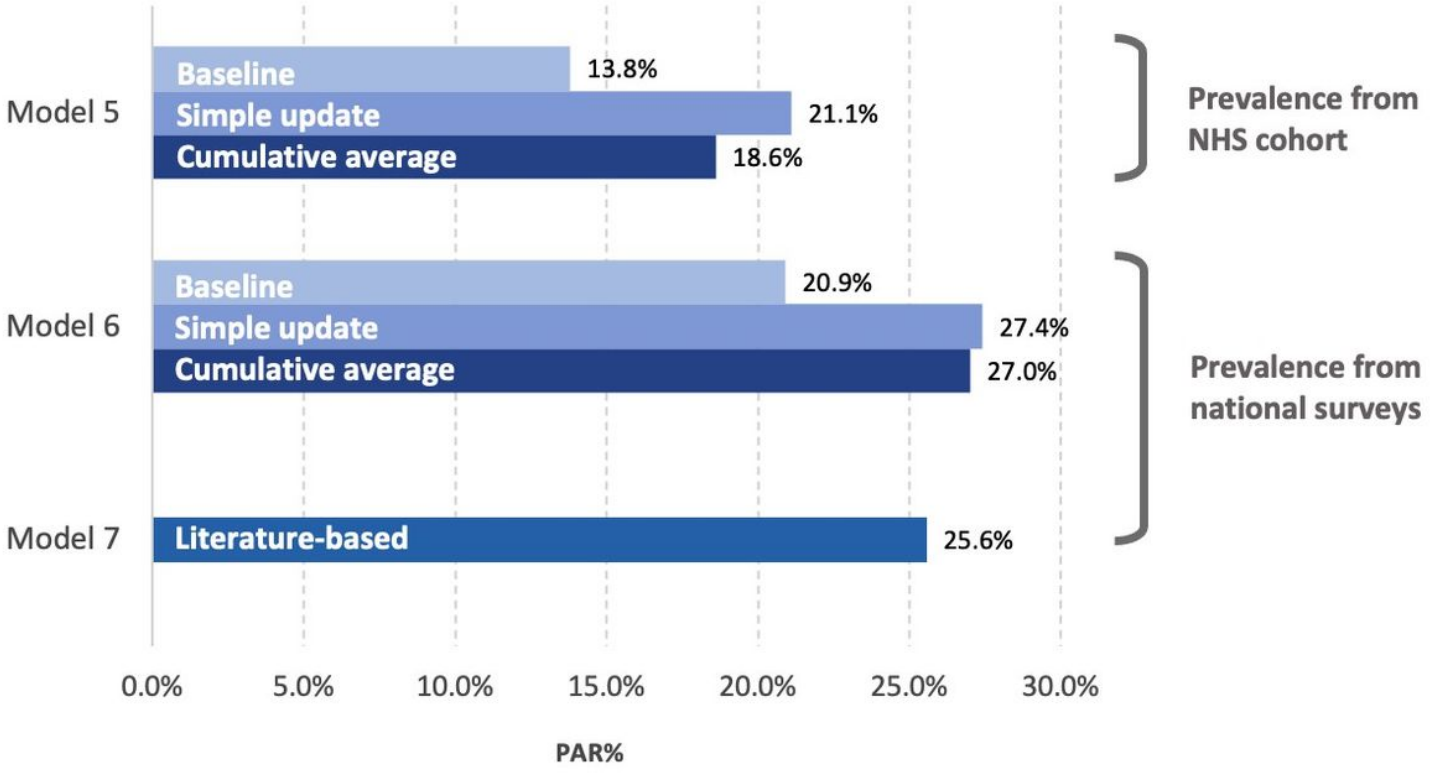

Figure 3

PAR% of postmenopausal breast cancer by alcohol, body mass index, and fruit and vegetable intake, and physical activity in the Nurses' Health Study and in the US by the conventional method

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 220614supplementaryinformation.docx