Calculation of population attributable risk for alcohol and breast cancer (United States)

Marilyn Tseng¹, Clarice R. Weinberg², David M. Umbach² & Matthew P. Longnecker¹,*
¹National Institute of Environmental Health Sciences Epidemiology Branch, Research Triangle Park, P.O. Box 12233 MD A3-05, Research Triangle Park, NC 27709, USA; ²National Institute of Environmental Health Sciences Biostatistics Branch, Research Triangle Park, NC, USA

Key words: Alcohol, breast-neoplasms, population attributable risk.

Abstract

Objectives: Because of increasing evidence that alcohol may be causally associated with breast cancer, we reconsider the population attributable risk (PAR) for alcohol and breast cancer for the US adult female population using an effect estimate from a meta-analysis and incorporating a revised perspective on measurement error correction.

Methods: To estimate PAR, we employed a formula appropriate to use with an adjusted effect estimate. To estimate intermediate quantities needed to apply that formula, we used adjusted relative risk estimates from a previously published meta-analysis, as well as SEER cancer statistics and general population data from the third National Health and Nutrition Examination Survey. We used relative risk estimates uncorrected for measurement error.

Results: The estimated age-adjusted PAR for alcohol and breast cancer was 2.1%.

Conclusions: Because of the modest association between alcohol and breast cancer and the generally moderate level of alcohol intake among US women, the proportion of breast cancer attributable to alcohol intake is small. Widespread efforts to reduce alcohol consumption would not have a substantial impact on breast cancer rates in this population. While selected subgroups of women might benefit from decreasing alcohol consumption, specific profiles for such women have yet to be defined and defended.

Introduction

A recent pooled analysis of data on alcohol and breast cancer [1] provides additional support for an association of alcoholic beverage consumption with increased risk of breast cancer in women. Because of the increasing evidence that alcohol may be causally linked to breast cancer, we reconsider the proportion of breast cancer attributable to alcohol consumption. Earlier work estimated the population attributable risk (PAR) to be 4 percent [2], but we will argue that this estimate is off by a factor of two. Two methodological issues render a re-estimation of the PAR for alcohol and breast cancer appropriate. First, many formulae for PAR in frequent use may not be appropriate to use with meta-analytic effect estimates because such estimates are often adjusted for confounding factors. The meta-analysis by Longnecker [2], for example, estimated the effect of alcohol on breast cancer risk adjusted, at the minimum, for age as a confounder. Second, a recent large study of breast cancer and alcohol [1] indicated that correcting for measurement error resulted in little change in the relative risk estimate, contrary to an earlier analysis [3]. This new result suggests that measurement error adjustment may not necessarily lead to more accurate estimation in this context. Here, we re-estimate PAR for alcohol and breast cancer for the adult (aged 20+) female population of the US based on results from a meta-analysis. We take into account the two methodological issues mentioned above: use of a formula for PAR that is more appropriate to use with a meta-analytic effect estimate, and a revised perspective on correcting alcohol effects for measurement error.
Materials and methods

We estimated PAR for four \((k = 3)\) levels of alcohol exposure: non-drinkers \((0 \text{ g/day})\), light drinkers \((\leq \text{half a drink per day, 0.1–6.4 g/day})\), moderate drinkers \((\text{half a drink to less than 2 drinks per day, 6.5–25.9 g/day})\), and heavy drinkers \((\geq 2 \text{ drinks per day, 26+ g/day})\). We considered age as the confounder of interest, with 14 \((l = 13)\) five-year age groups \((20–24, 25–29, \ldots, 80–84, 85+)\).

To estimate the PAR for alcohol and breast cancer using an adjusted RR derived from meta-analysis, we used a formula described by Bruzzi et al. \([4]\), for an exposure with \(k + 1\) levels:

\[
\text{PAR} = 1 - \sum_{i=0}^{k} \frac{p_{(c)i}}{RR_i}
\]

\(p_{(c)i}\) is the proportion of cases at the \(i\)-th level of exposure, and \(RR_i\) represents the confounder-adjusted risk of the outcome at the \(i\)-th level of exposure relative to those with \(i = 0\).

\(RR_i\) was calculated as \(\exp[\beta_{\text{alcohol}} \cdot x_i]\) where \(\beta_{\text{alcohol}}\) represents the log relative risk per unit change in g/day of ethanol intake and \(x_i\) is the estimated median intake for each category of alcohol consumption. For \(\beta_{\text{alcohol}}\) we used an estimate of 0.0076 \((SE = 0.0017)\), the result of a meta-analysis including US studies only \([2]\). Median levels of alcohol consumption for non-, light, moderate, and heavy drinkers were estimated for the adult female population based on data from the third National Health and Nutrition Examination Survey (NHANES III) \([5]\). Participants in the NHANES III interview who reported drinking at least 12 alcoholic drinks during the past year were asked how many days they drank during the last year, and on average how many drinks they consumed on each day they drank. Amount of ethanol in grams consumed per day was calculated as:

\[
\text{Amount of ethanol in grams consumed per day} = (\# \text{ days drank alcohol}) \times \text{(average \# drinks/day)} \times (13 \text{ g ethanol/drink}).
\]

Women who had not had at least 12 drinks during the past year were categorized as non-drinkers.

While the estimate for \(RR_i\) was derived from meta-analysis, calculating the value of \(p_{(c)i}\) for equation 1 was more complex. When data are available for the joint distribution of the exposure and its confounders in the general population, the proportion of cases exposed for a given level \(i\) of alcohol intake, \(p_{(c)i}\), can be estimated by taking into account the distribution of a confounder with \(l + 1\) levels. Here, \(p_{ij}\) is the proportion of women in alcohol category \(i\) and age group \(j\) in the general population, and \(RR_j\) represents risk for those in that alcohol-age category relative to those in a referent \((i = 0, j = 0)\) category.

For equation 2, information on the joint distribution of alcohol consumption and age in the adult female US population, \(p_{ij}\), was obtained from NHANES III \([5]\). A source for \(RR_j\) was less straightforward. If the simplifying assumption is made that there is no interaction on the multiplicative scale, then \(RR_j\) can be estimated as \(RR_{ij} = RR_i \cdot RR_j\), so that all that is required are estimates of the confounder-adjusted exposure effect \((RR_i)\) and the exposure-adjusted confounder effect \((RR_j)\).

As mentioned, the age-adjusted alcohol effect, \(RR_i\), was estimated via meta-analysis. To estimate the alcohol-adjusted age effect, \(RR_{ij}\), we adjusted SEER 1990–1994 age-specific incidence rates \((R(w)_j)\) of breast cancer \([6]\) for alcohol intake by the following algebraic strategy. The incidence rate for age group \(j\) represents a weighted average of breast cancer rates across alcohol categories, or:

\[
R_{(w)j} = \frac{1}{p_j} \sum_{i=0}^{k} p_{ij} \cdot R_{ij} = \frac{R_{0j}}{p_j} \sum_{i=0}^{k} p_{ij} \cdot \frac{R_{ij}}{R_{0j}}
\]

where \(p_j = \sum_{i=0}^{k} p_{ij}\), or the total proportion of the population in age group \(j\) from general population data. Since by assumption \(\frac{R_{0j}}{R_{0j}} = RR_i\) for any value of \(j\), then substituting for \(\frac{R_{0j}}{R_{0j}}\) in the rightmost member of equation 3 and solving the resulting expression for \(R_{0j}\), the incidence rate for nondrinkers \((i = 0)\) in age group \(j\) adjusted for alcohol intake becomes:

\[
R_{0j} = \frac{R_{(w)j} \cdot p_j}{\sum_{i=0}^{k} p_{ij} \cdot RR_i}.
\]

Under the assumption of no interaction, \(RR_j = \frac{R_j}{R_0}\) for any value of \(j\). In particular, taking \(i = 0\), \(RR_j = \frac{R_j}{R_0}\).

Substituting equation 4 into both the numerator and the denominator of this last expression (where \(j\) is set to 0 in the denominator) yields:

\[
RR_{ij} = \frac{R_{(w)j} \cdot p_j}{R_{(w)0} \cdot p_0} \left[ \sum_{i=0}^{k} p_{ij} \cdot RR_i \right] \left[ \sum_{i=0}^{k} p_{ij} \cdot RR_i \right]^{-1}
\]
In this way, we used the unadjusted population incidence rates to estimate relative risks for age now adjusted for alcohol.

### Results

Median levels of alcohol consumption for non-, light, moderate, and heavy drinkers in the adult female population of NHANES III were 0 g, 2.0 g, 11.1 g, and 33.3 g ethanol per day, respectively [5]. Based on a $\beta_{\text{alcohol}}$ of 0.0076 [2], the resulting $RR_{ij}$ that we used to estimate PAR using equation 1 was 1.0 for non-drinkers, 1.0 for light drinkers, 1.1 for moderate drinkers, and 1.3 for heavy drinkers. Estimates for $RR_{ij}$ are shown in Table 1, and estimates for $p_{ij}$, obtained from NHANES III [5], are shown in Table 2. When those values were used in equation 2, the proportions of cases in each category of alcohol consumption ($p_{ij}x$) were 67.1% non-drinkers, 17.7% light drinkers, 11.2% moderate drinkers, and 4.1% heavy drinkers. The resulting estimate of the age-adjusted PAR for alcohol and breast cancer was 2.1%.

Further adjustment for factors in addition to age would not materially change estimates of the alcohol parameter (Stephanie A. Smith-Warner, personal communication, 1998). Although methods are available to calculate variance of PAR estimated using other approaches [7, 8], a method is not yet readily available to calculate variance of PAR estimated as described above. To examine the sensitivity of our PAR estimate to the magnitude of the meta-analytic estimate used for $\beta_{\text{alcohol}}$, we recalculated PAR using the lower and upper

### Table 1. Relative risk for each alcohol x age category ($RR_{ij}$)

<table>
<thead>
<tr>
<th>Age group (years)</th>
<th>Alcohol category</th>
<th>Non-drinker (0 g/day)</th>
<th>Light drinker (0.1-6.4 g/day)</th>
<th>Moderate drinker (6.5-25.9 g/day)</th>
<th>Heavy drinker (26+ g/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-24</td>
<td>1.0</td>
<td>1.0</td>
<td>1.1</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>25-29</td>
<td>5.5</td>
<td>5.6</td>
<td>6.0</td>
<td>7.1</td>
<td></td>
</tr>
<tr>
<td>30-34</td>
<td>19.3</td>
<td>19.5</td>
<td>21.0</td>
<td>24.8</td>
<td></td>
</tr>
<tr>
<td>35-39</td>
<td>48.1</td>
<td>48.7</td>
<td>52.3</td>
<td>61.9</td>
<td></td>
</tr>
<tr>
<td>40-44</td>
<td>94.0</td>
<td>95.4</td>
<td>102.3</td>
<td>121.1</td>
<td></td>
</tr>
<tr>
<td>45-49</td>
<td>153.9</td>
<td>156.1</td>
<td>167.5</td>
<td>198.3</td>
<td></td>
</tr>
<tr>
<td>50-54</td>
<td>188.8</td>
<td>191.5</td>
<td>205.5</td>
<td>243.3</td>
<td></td>
</tr>
<tr>
<td>55-59</td>
<td>218.8</td>
<td>221.9</td>
<td>238.1</td>
<td>281.9</td>
<td></td>
</tr>
<tr>
<td>60-64</td>
<td>266.0</td>
<td>269.8</td>
<td>289.4</td>
<td>342.7</td>
<td></td>
</tr>
<tr>
<td>65-69</td>
<td>314.7</td>
<td>319.2</td>
<td>342.5</td>
<td>405.5</td>
<td></td>
</tr>
<tr>
<td>70-74</td>
<td>353.7</td>
<td>358.7</td>
<td>384.9</td>
<td>455.7</td>
<td></td>
</tr>
<tr>
<td>75-79</td>
<td>372.7</td>
<td>378.0</td>
<td>405.6</td>
<td>480.2</td>
<td></td>
</tr>
<tr>
<td>80-84</td>
<td>366.2</td>
<td>371.4</td>
<td>398.4</td>
<td>471.8</td>
<td></td>
</tr>
<tr>
<td>85+</td>
<td>332.6</td>
<td>337.4</td>
<td>362.0</td>
<td>428.6</td>
<td></td>
</tr>
</tbody>
</table>

95% confidence bounds for $\beta_{\text{alcohol}}$ from meta-analysis. This produced PAR estimates that ranged from 1.2% to 2.9%. We also recalculated PAR using a distribution of alcohol intake resembling that of the Nurses’ Cohort. Based on data presented by Fuchs et al. [9], approximate proportions of cases in each category of alcohol consumption were 31% non-, 26% light, 33% moderate, and 10% heavy drinkers. The resulting PAR based on such a distribution was 7.4%. The remarkable difference in PAR highlights an important fact: PAR can differ substantially across subpopulations that have similar relative risks but strikingly different distributions of alcohol consumption.

### Discussion

We estimated the age-adjusted PAR for alcohol and breast cancer to be 2.1%. Our estimate of PAR relied on the assumption that alcohol consumption is linearly associated with the logit of breast cancer risk. The meta-analysis by Longnecker [2] found evidence of a log-linear dose-response between alcohol and breast cancer. Also, in their pooled analysis, Smith-Warner et al. [1] found a logit-linear increase in breast cancer risk for alcohol intake less than 60 g/day, after which there was little increase in risk with increasing intake. As less than 0.5% of the women in NHANES III consumed at least 60 g of alcohol a day, accounting for this flattening of the risk curve is unlikely to change the PAR estimate.
meaningfully. We also assumed that alcohol and age have joint effects that are multiplicative, and previous research gives little reason to expect departure from this model.

In our calculation of the PAR for alcohol and breast cancer, we used an estimate of the alcohol effect that was uncorrected for measurement error. Use of an uncorrected estimate is the primary source of the discrepancy between the present PAR estimate, 2.1%, and a previous estimate of 4% [2] based on a $\beta_{\text{alcohol}}$ that was enlarged by correcting for measurement error [3]. In a recent, large, pooled analysis, however, the overall impact on the estimated alcohol effect of adjusting for measurement error was negligible, although confidence limits, of course, broadened [1]. This result suggests that error correction may not be crucial even though reported alcohol consumption is subject to both systematic and random measurement errors.

That an uncorrected estimate might remain relatively unbiased in the presence of strong measurement error is plausible for alcohol and breast cancer because the effects of bias and random error may counteract each other. In general, when measurement of a risk factor is unbiased (i.e., without systematic error) but subject to random error, the random error alone causes the odds ratio estimated using the measured covariate to be closer to one on average than the true odds ratio. As most epidemiologists appreciate, this attenuation becomes more extreme as the errors worsen [10]. For example, if random error reduces the intraclass correlation among repeated measures of alcohol intake to 0.80 [11], then the logarithm of the observed alcohol odds ratio would be on average 20% too small [12].

On the other hand, systematic within-person measurement error can introduce bias into a measured risk factor and into the resulting estimate of its effect even without random measurement error, and systematic underreporting will tend to inflate the risk estimate [13]. Under the measurement error model of Rosner, Willett, and Spiegelman [13] for continuous risk factors, for example, if respondents report only 75% of their consumption, then the observed dose-response slope can be on average a third greater than the actual slope. Similar inflation can be seen with polychotomous predictors [14]. With alcohol, widespread underreporting is likely. According to national survey data, the average reported daily intake of ethanol among US adults is 14 g [15]—substantially less than the adult per capita intake of 21 g/day estimated from sales data [16]. Thus, up to one-third of the alcohol consumed in the US is unaccounted for [17]. Because heavy drinkers are underrepresented in surveys [18], the proportion of alcohol intake reported among respondents may be higher than two-thirds. Also, the dose-response might still be attenuated if heavy drinkers report no consumption at all. Nevertheless, the main point remains that underreporting can inflate the alcohol-disease association beyond its true value.

Whenever random and systematic within-person measurement error are both important, they compete to influence the apparent alcohol effect estimated from reported intake. Systematic underreporting will tend to offset to an unknown extent the attenuating influence of random error so that bias in the relative risk estimates based on reported intake may be small. Thus, basing PAR estimates on relative risk estimates without correcting for measurement error seems justified. Of course, the nature of the relationship between self-reported alcohol intake and true intake is complex. In the future, a more detailed understanding of this relationship may lead to better measurement error models for correcting alcohol relative risk estimates and prompt us to re-evaluate our use of uncorrected estimates.

Whether the effect estimate we used for alcohol in the calculations above represents the most informative aspect of exposure is another consideration. For example, while the meta-analytic effect estimate used in these calculations was for recent alcohol intake [2], some evidence suggests that lifetime intake may be more informative [19]. Because this suggestion has not been established, however, the present calculation, based on recent alcohol use, provides a reasonable estimate of effect to use in calculating the PAR.

The estimated proportion of breast cancer attributable to alcohol use, 2%, is small because of the modest association between alcohol and breast cancer, and the generally moderate level of alcohol intake among US women, although alcohol consumption may be higher in certain subpopulations. A widespread effort to reduce alcohol consumption would not have a substantial impact on breast cancer rates in the US, and might in fact increase overall mortality because of effects on cardiovascular disease [9]. Selected subgroups of women, for example, those at low risk of heart disease and high risk of breast cancer, might benefit from decreasing alcohol consumption, but specific profiles for such women have yet to be defined and defended.

**Acknowledgements**

The authors thank Dr Glinda Cooper for her valuable comments on this manuscript.
References


