Coffee Drinking and Risk of Lung Cancer—A Meta-Analysis

Vania Galarraga and Paolo Boffetta

Abstract

Background: Previous epidemiologic results on coffee consumption and lung cancer risk have not been consistent. Furthermore, not all studies have addressed the potential role of tobacco as a confounder in this association. A meta-analysis was conducted to assess the effect of coffee consumption on lung cancer risk independent of tobacco use.

Methods: A systematic review and a meta-analysis based on random effects models were performed using studies from the PubMed and EMBASE databases, and the references from the retrieved articles. Included were 8 prospective cohorts and 13 case-control studies, which provided data for 19,892 cases and 623,645 non-cases, timeframe 1986-2015.

Results: The meta-relative risk (RR) for coffee drinking, not controlling for tobacco smoking, was 1.09 [95% confidence interval (CI), 1.00–1.19], the reference group was never drinkers. There was significant heterogeneity among the study results (Q = 84.39, I² = 75.1%, P heterogeneity < 0.001). Among non-smokers, coffee was not associated with lung cancer risk (RR, 0.92; 95% CI, 0.75–1.10), the reference group was never drinkers. The meta-RR for 1 cup per day increase, unadjusted for smoking, was 1.04 (95% CI, 1.03–1.05); the corresponding RR for non-smokers was 0.95 (95% CI, 0.83–1.09).

Conclusions: The pooled estimates indicated that when the potential confounding effect from smoking is controlled for, coffee drinking does not appear to be a lung cancer risk factor. Further pooled analyses, with larger non-smokers population size, are encouraged to confirm these results.

Impact: This study illustrates that the association between coffee consumption and lung cancer can be confounded by tobacco smoking. Cancer Epidemiol Biomarkers Prev; 25(6): 951-7. ©2016 AACR.

Introduction

Lung cancer is the most common cancer in the world with incidence and mortality rates higher than any other neoplasm (1). Because of late-stage diagnosis, such malignant growths are frequently hard to treat effectively and can easily metastasize if left untreated. There are well-established risk factors for lung cancer, such as tobacco smoking, selected occupational agents, radon decay products, and indoor and outdoor air pollution. Dietary factors such as vegetables and fruits have been suggested to provide a protective effect (2).

Previous epidemiologic studies have evaluated the potential association between coffee consumption and risk of lung cancer, but the results have not been consistent. An important aspect to consider is the potential confounding effect from tobacco smoking, a known cause of lung cancer, which in many populations is associated with coffee drinking (3). Although several studies have attempted to control for the confounding effect of tobacco smoking via statistical adjustment, residual confounding is difficult to rule out completely. In a meta-analysis by Tang and colleagues (4) results were provided that indicated a significant positive association between coffee intake and lung cancer, but the results of the individual studies were inconsistent, and the authors did not take into account the possible residual confounding effect from smoking. A recently published meta-analysis by Xie and colleagues (5) provides results indicating an association between coffee consumption and risk of lung cancer; however, this meta-analysis included a limited number of studies.

An association between coffee drinking and lung cancer risk is justified by the fact that coffee contains agents which may cause cancer under experimental conditions, such as acrylamide (6), which is formed at very low levels during the roasting of coffee beans. In contrast, other agents present in coffee have been reported to exert an anticarcinogenic effect (7), including the diterpenes cafestol and kahweol (8). Different mechanisms appear to be involved in these chemoprotective effects, such as induction of conjugating enzymes (e.g., GST-glutathione S-transferases), increased expression of proteins involved in cellular antioxidant defense (e.g., g-glutamyl cysteine synthetase), and inhibition of the expression or activity of cytochromes P450 involved in carcinogen activation. All of these have been hypothesized to act through an inhibition of the formation or the stimulation of the detoxification of carcinogen intermediates, resulting in decreased DNA damage and in the blocking of tumor formation (7). Furthermore, there is strong evidence that the risk of several human cancer is either reduced (e.g., liver, endometrium, colorectal, prostate cancer; refs. 9, 10) or unaffected (e.g., stomach, breast, ovarian cancer; ref. 10) in coffee drinkers compared with non-drinkers, which reduces the plausibility of a positive association with lung cancer.
The objective of this systematic review and meta-analysis was to assess the association between coffee consumption and lung cancer independent of tobacco use, in particular among never smokers.

**Materials and Methods**

**Search strategy**

Systematic search of two electronic databases PubMed, and EMBASE was performed for relevant papers. The key terms and MESH terms used for the search on coffee, diet, dietary factors, food habits, and lung cancer are provided in the Supplementary Information Section (Supplementary Materials and Methods). The titles and abstracts were examined as part of the preliminary screening process, and full texts were obtained for those articles that contained information within the abstracts about coffee (or dietary factors) and lung cancer. The lists of references of the relevant articles were reviewed with the goal of identifying additional studies.

The total number of articles identified in the literature search was 1,493; from these, relevant data were identified and extracted from 21 articles (2, 11–30) after the selection criteria were considered. Further details on the selection process are provided in Fig. 1, and specific characteristics for the selected studies are provided in Table 1.

**Study selection**

The studies included in the meta-analysis comprised adult patients with lung cancer, and control (unaffected) individuals. All histologic types were included, diagnosis of the study subjects was ascertained via histology for some studies (2, 21, 27, 28), whereas the rest had general lung cancer diagnosis without

---

**Figure 1.**

Presents the study selection schematic for how the records in the analysis were identified.
specifying histologic type. Studies were either of prospective (cohort) or retrospective (case–control) design, and had to report original data.

Articles were excluded if they were not related to coffee and lung cancer or if they did not provide a measure of association and its SE (or confidence interval) between lung cancer and coffee consumption (or sufficient raw data to calculate them). Also excluded were meta- and pooled analyses with comprehensive data from previous publications to avoid duplicated data. For studies that had more than one report, the most recent report with the most comprehensive information was used.

Studies that were not in English, Spanish, or French were excluded a priori. However, we only encountered 3 studies that were not published in these languages one in Chinese (31), one in Danish (32), and one in Polish (33); furthermore, these did not appear to have sufficiently detailed data to be included.

### Data extraction

The primary outcome of interest was lung cancer and this was measured as a dichotomous variable. The exposure of interest was coffee consumption, which was measured as a continuous variable (cups/day), and as a dichotomous variable (ever/never drinkers). Tobacco smoking status was the primary potential confounder, and was measured differently in the different studies. Additional potential confounders included education, occupational exposures, and history of non-malignant respiratory disease (tuberculosis or chronic obstructive pulmonary disease). We extracted data on relative risks (RR) or ORs with corresponding 95% confidence intervals (CI). Lung cancer incidence data were used in all but one study (15), which was based on mortality.

### Statistical analysis

The measure of association of interest was the RR for prospective cohort studies, and the OR for the case–controls studies, with corresponding 95% CIs. Never coffee drinkers comprised the reference (unexposed) category. An overall pooled RR was estimated, together with its 95% CI, based on the individual estimates from each study. Each study estimate of the RR was given a weight based on the inverse of the variance of the effect estimate.

A random effects model was used because of the heterogeneity in the design and analysis of the available studies (34). Methods that quantify inconsistency across studies have been developed (35), not only to determine whether heterogeneity is present but also to assess its impact on the meta-analysis. The I^2 statistic, which describes the percentage of variation across studies that is due to heterogeneity rather than chance, was used (35).

Meta-regression analysis was done to obtain study specific dose–response (11–13, 15, 17, 21–24, 26–30) based on data extracted at different frequencies of coffee consumption (cups/day). The RR for the dose–response was obtained and plotted against an established common scale of 1 cup per day. The study-specific slopes were derived from the natural logarithm of the risk estimates across the studies. These analyses were executed with the commands metan and metareg of the statistical software STATA Version 11.1 (StataCorp.).

In meta-analyses, bias may arise from publication bias, where articles that do not show positive associations of increased risk are often not published. To assess whether our search was subject to publication bias, the Egger's test (36) was performed.

### Table 1. Key aspects of studies included in the meta-analysis

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study design</th>
<th>Sex</th>
<th>Country</th>
<th>Cases (n)</th>
<th>Prevalence of coffee drinkers (controls) %</th>
<th>Results</th>
<th>Smoking status</th>
<th>Adjusted for smoking</th>
<th>RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guerin (28)</td>
<td>Prospective</td>
<td>M-F</td>
<td>USA</td>
<td>9,196</td>
<td>87.85</td>
<td>FE</td>
<td>All</td>
<td>Y</td>
<td>1.10 (1.05–1.15)</td>
</tr>
<tr>
<td>Hashibe (29)</td>
<td>Prospective</td>
<td>M-F</td>
<td>USA</td>
<td>1,337</td>
<td>66.53</td>
<td>FE</td>
<td>All</td>
<td>Y</td>
<td>1.04 (1.01–1.07)</td>
</tr>
<tr>
<td>Saninik (30)</td>
<td>PCC</td>
<td>M-F</td>
<td>France</td>
<td>2,684</td>
<td>94.28</td>
<td>PR</td>
<td>All</td>
<td>Y</td>
<td>1.08 (0.80–1.47)</td>
</tr>
<tr>
<td>Luqm (20)</td>
<td>PCC</td>
<td>M-F</td>
<td>Pakistan</td>
<td>400</td>
<td>5.25</td>
<td>PR</td>
<td>All</td>
<td>Not specified</td>
<td>1.80 (1.10–2.80)</td>
</tr>
<tr>
<td>Bae (12)</td>
<td>Prospective</td>
<td>M</td>
<td>Korea</td>
<td>93</td>
<td>4.20</td>
<td>FE</td>
<td>Smokers</td>
<td>N</td>
<td>1.61 (0.85–3.02)</td>
</tr>
<tr>
<td>Ganesh (16)</td>
<td>HCC</td>
<td>M</td>
<td>India</td>
<td>408</td>
<td>15.33</td>
<td>PR</td>
<td>All</td>
<td>Not specified</td>
<td>2.90 (2.20–3.70)</td>
</tr>
<tr>
<td>Chou (14)</td>
<td>HCC</td>
<td>F</td>
<td>China</td>
<td>279</td>
<td>23.30</td>
<td>PR</td>
<td>All</td>
<td>Y</td>
<td>0.75 (0.48–1.16)</td>
</tr>
<tr>
<td>Kubik (2)</td>
<td>HCC</td>
<td>M-F</td>
<td>Czech Republic</td>
<td>1,096</td>
<td>2.80</td>
<td>FE</td>
<td>All</td>
<td>Y</td>
<td>0.87 (0.68–1.11)</td>
</tr>
<tr>
<td>Baker (13)</td>
<td>HCC</td>
<td>M-F</td>
<td>USA</td>
<td>993</td>
<td>74.74</td>
<td>FE</td>
<td>All</td>
<td>Y</td>
<td>1.30 (1.08–1.56)</td>
</tr>
<tr>
<td>Khan (19)</td>
<td>Prospective</td>
<td>M-F</td>
<td>Japan</td>
<td>51</td>
<td>78.14</td>
<td>FE</td>
<td>All</td>
<td>Y</td>
<td>0.80 (0.50–1.60)</td>
</tr>
<tr>
<td>Hu (17)</td>
<td>PCC</td>
<td>F</td>
<td>Canada</td>
<td>161</td>
<td>74.95</td>
<td>FE</td>
<td>Non-smokers</td>
<td>N/A</td>
<td>0.90 (0.60–1.32)</td>
</tr>
<tr>
<td>Takezaki (27)</td>
<td>HCC</td>
<td>M-F</td>
<td>Japan</td>
<td>1,045</td>
<td>2.60</td>
<td>FE</td>
<td>All</td>
<td>Y</td>
<td>0.99 (0.88–113)</td>
</tr>
<tr>
<td>Mendillahasu (21)</td>
<td>HCC</td>
<td>M</td>
<td>Uruguay</td>
<td>427</td>
<td>54.67</td>
<td>PR</td>
<td>Smokers</td>
<td>Y</td>
<td>1.11 (0.72–1.73)</td>
</tr>
<tr>
<td>Nyberg (24)</td>
<td>PCC</td>
<td>M-F</td>
<td>Sweden</td>
<td>124</td>
<td>90.21</td>
<td>PR</td>
<td>Non-smokers</td>
<td>N/A</td>
<td>0.57 (0.27–1.22)</td>
</tr>
<tr>
<td>Axelsson (11)</td>
<td>PCC</td>
<td>M</td>
<td>Sweden</td>
<td>308</td>
<td>92.00</td>
<td>PR</td>
<td>All</td>
<td>Y</td>
<td>0.94 (0.38–2.29)</td>
</tr>
<tr>
<td>Sankaranarayanan (25)</td>
<td>HCC</td>
<td>M</td>
<td>India</td>
<td>281</td>
<td>45.65</td>
<td>PR</td>
<td>All</td>
<td>Y</td>
<td>0.69 (0.48–1.07)</td>
</tr>
<tr>
<td>Stensvold (26)</td>
<td>Prospective</td>
<td>M-F</td>
<td>Norway</td>
<td>162</td>
<td>2.20</td>
<td>FE</td>
<td>All</td>
<td>Y</td>
<td>1.90 (1.40–2.70)</td>
</tr>
<tr>
<td>Chow (15)</td>
<td>Prospective</td>
<td>M</td>
<td>USA</td>
<td>219</td>
<td>72.93</td>
<td>FE</td>
<td>All</td>
<td>Y</td>
<td>2.20 (1.70–2.80)</td>
</tr>
<tr>
<td>Metllin (22)</td>
<td>HCC</td>
<td>M-F</td>
<td>USA</td>
<td>569</td>
<td>72.93</td>
<td>FE</td>
<td>All</td>
<td>Y</td>
<td>1.06 (0.85–1.33)</td>
</tr>
<tr>
<td>Nomura (23)</td>
<td>Prospective</td>
<td>M</td>
<td>Japan</td>
<td>110</td>
<td>84.10</td>
<td>FE</td>
<td>All</td>
<td>N</td>
<td>1.55 (0.85–2.81)</td>
</tr>
<tr>
<td>Jacobsen (18)</td>
<td>Prospective</td>
<td>M-F</td>
<td>Norway</td>
<td>177</td>
<td>74.30</td>
<td>FE</td>
<td>All</td>
<td>N</td>
<td>1.82 (1.27–2.61)</td>
</tr>
</tbody>
</table>

NOTE: All includes never, former, and current smokers. Smokers include current and former smokers.

Abbreviations: FE, calculated via fixed-effect model; PR, used published results.

*Data on mortality.

*Prevalence of coffee drinkers in full cohort.
Results

From the 21 studies included in the meta-analysis, 8 were of prospective cohort design and 13 of case-control design (Supplementary Table S1); these studies included a total of 19,892 cases and 623,645 non-cases. To obtain comparable results across studies, we first conducted a number of within-study meta-analyses, based on fixed-effect models, to estimate summary results, for example, combining gender-specific results if an overall risk estimate was not reported. The results of these within-study meta-analyses are reported in Table 1.

Pooled analysis using the random effects model was conducted for risk of lung cancer and coffee consumption as a dichotomous variable (ever vs. never drinkers). As provided in Table 2, the results showed a RR of 1.09 (95% CI, 1.00–1.19) for those who consumed coffee when compared with the reference group, those who did not consume coffee, without controlling for the effect due to smoking. Figure 2 depicts the forest plot of the pooled results from the 21 studies. There was significant heterogeneity among the study results ($Q = 84.39, I^2 = 75.1\%$, $P_{\text{heterogeneity}} < 0.001$). After restricting the results to the studies that adjusted for smoking, the RR was 1.03 (95% CI, 0.95–1.12) from Table 2.

Table 2. Pooled analysis using random effects model for ever versus never coffee drinking, stratified by smoking status

<table>
<thead>
<tr>
<th>Smoking status</th>
<th>Sex</th>
<th>No. studies</th>
<th>RR (95% CI)</th>
<th>$P_{\text{heterogeneity}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All*a</td>
<td>M-F</td>
<td>21</td>
<td>1.09 (1.00–1.19)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>All*a</td>
<td>M</td>
<td>13</td>
<td>1.31 (1.02–1.59)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>All*a</td>
<td>F</td>
<td>7</td>
<td>0.87 (0.73–1.00)</td>
<td>0.459</td>
</tr>
<tr>
<td>Never smokers</td>
<td>M-F</td>
<td>4</td>
<td>0.92 (0.75–1.10)</td>
<td>0.244</td>
</tr>
<tr>
<td>Smoking adjusted*b</td>
<td>M-F</td>
<td>16</td>
<td>1.03 (0.95–1.12)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Smoking unadjusted*b</td>
<td>M-F</td>
<td>6</td>
<td>1.79 (1.23–2.34)</td>
<td>&lt;0.010</td>
</tr>
</tbody>
</table>

*aAll includes the never, former, and current smokers.

*bSmoking adjusted, and unadjusted, for the current and former smokers.

Due to smoking. Figure 2 depicts the forest plot with the individual, and pooled, RRs from the 21 studies in the analysis.
heterogeneity among these results was lower \((Q = 51.02, I^2 = 70.6\%; P_{\text{heterogeneity}} < 0.001)\). The forest plot of the pooled estimate for the studies that adjusted for smoking is provided in Supplementary Fig. S1. In studies of only non-smokers, coffee was not associated with the risk for lung cancer \((RR, 0.92; 95\% \text{ CI}, 0.75–1.10)\), the heterogeneity was very small \((Q = 4.5, I^2 = 28\%)\).

The forest plot (Supplementary Fig. S2) with the pooled results from the non-smokers studies provides a graphical representation of the results for these studies.

Meta-analyses were also executed according to study design, hospital-based case–control (HCC), population-based case–control (PCC), and prospective cohort, stratifying for smoking status. Studies of prospective cohort design showed an increased risk in the group where smoking was controlled for \((1.14; 95\% \text{ CI}, 1.01–1.28; P_{\text{heterogeneity}} < 0.001)\), Supplementary Table S2.

We further examined the association of coffee and lung cancer risk stratifying by region where the study was conducted, in addition to smoking status. We were able to extract data from studies corresponding to four different continents: Asia, Europe, North America, and South America. The results of the meta-analysis were not heterogeneous across geographic regions, Supplementary Table S3.

Pooled analysis using random-effects model for ever versus never coffee consumption and lung cancer risk was done for studies, which also controlled for occupational exposures and education. For the studies that adjusted for occupational education, the RR was 1.05 \((95\% \text{ CI}, 0.95–1.16; P_{\text{heterogeneity}} < 0.001)\). The studies that did not adjust for neither occupation nor education, which were \((11–13, 18, 19, 21, 23, 24, 26)\), resulted in a RR of 1.21 \((95\% \text{ CI}, 0.95–1.48; P_{\text{heterogeneity}} = 0.02)\).

Dose–response analysis was carried out for the 21 studies, and risk was estimated for lung cancer for increase in 1 cup per day of coffee consumed. There was an observed dose–response with RR of 1.04 for an increase of 1 cup per day \((95\% \text{ CI}, 1.03–1.05)\), not adjusting for smoking (Fig. 3). The dose–response analysis from the results in never-smokers \((N = 4)\), resulted in a RR of 0.95 \((95\% \text{ CI}, 0.83–1.09; \text{Fig. 4})\). When the dose–response analysis was conducted on the studies that adjusted for smoking \((N = 13)\), the results were similar to those reported in Fig. 3 \((RR, 1.04; 95\% \text{ CI}, 1.03–1.05)\).

The results of the Egger’s test provided evidence of publication bias due to the lacking of small published studies with negative results \((bias, 1.11; 95\% \text{ CI}, 0.72–1.49; P < 0.001)\). A funnel plot was also generated (Supplementary Fig. S3) to provide visual representation of how publication bias skews the results.

**Discussion**

This meta-analysis was comprised of 8 prospective cohort studies, 5 PCCs, and 8 HCCs; the analysis was performed to assess the effects of coffee consumption on lung cancer independent of tobacco use. The overall pooled estimates provided evidence that if the residual-confounding effect due to smoking is controlled, coffee does not appear to be a risk factor for lung cancer.

Results of observational studies of lung cancer, and their meta-analyses, may be confounded by tobacco smoking even if this factor is adjusted for in the statistical analysis. Misclassification of the confounder is a cause of incomplete adjustment, which results in residual confounding. Studies of non-smokers provide strong evidence of an effect of the suspected risk factors, independent from smoking. Estimating and assessing levels of residual confounding present in a study has been previously studied, and methods for quantifying the degree of confounding misclassification have been developed (37). As per the method presented by Savitz and colleagues (37), under the hypothesis that there was no association between coffee and risk of lung cancer, our smoking-adjusted results had 80% of the excess risk in the unadjusted analysis removed \([1.21–1.04]/1.21\], and one should consider whether the remaining 20% represents a real effect or if it is due to residual confounding because of misclassification of smoking. In fact, even a small amount of misclassification of smoking (95% of both sensitivity and specificity, assuming a conservative RR of 5 for the association between smoking and lung cancer), would produce the amount of residual confounding observed in our analysis. This level of sensitivity and specificity has been found

---

**Figure 3.**

Presents the dose–response graph not adjusted for smoking status for 1 cup of coffee per day. Note: Circles represent the risk estimates from each of the studies; the size of the circles is proportional to the SE. We did not estimate these risk estimates; they were collected from each individual study.
in studies of self-reported smoking habits (38). Furthermore, given the correlation between tobacco smoking and coffee drinking (3), residual confounding is likely to affect also the dose–response analysis, as confounding will be stronger for higher levels of coffee consumption.

An advantage of our systematic review was the thoroughly executed search with of all available published studies within our time limit, which had any potential information and data on this subject. We were able to include a relatively large number of studies, including the most recently published studies on the topic, and conduct several subgroup analyses using fixed-effect models. The use of meta-regression was employed to examine dose–response relationships.

A potential limitation in the systematic review may be that only studies published in English, Spanish, or French were considered. However, we only encountered 3 studies that were not published in these languages (31–33). The number of studies for non-smokers that provided data on the association between coffee and lung cancer risk was also limited. Additional studies, with larger population size, should be performed in the non-smokers population to further determine whether the inverse association in dose–response we found in this study represents a truly protective effect, as it has been shown for several other cancers (8, 9), or if it is simply due to random fluctuation. Our analyses were not stratified by use of caffeinated versus decaffeinated coffee because the data were not strong enough to address this aspect of the association between coffee intake and lung cancer. A further limitation was the lack of information on coffee preparation methods. Nevertheless, for the purposes of this meta-analysis, it was assumed that only the amount of water changed in the preparation.

Justification of the large statistically significant heterogeneity within the studies, for the pooled association of coffee and risk of lung cancer, can be attributed to sources such as study methodology, approaches to data acquisition based on the diversity of questionnaires, and follow-up periods.

Conclusions of our meta-analysis based on the pooled estimates, indicate that when the potential confounding effect from smoking is controlled for, coffee drinking does not constitute a risk factor for lung cancer. Additional pooled analyses, with larger population size in non-smokers are recommended to confirm these results.

Disclosure of Potential Conflicts of Interest
Paolo Boffetta has provided expert testimony for Morrison & Foerster. No potential conflicts of interest were disclosed by the other author.

Authors’ Contributions
Conception and design: P. Boffetta
Development of methodology: V. Galarraga, P. Boffetta
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): V. Galarraga, P. Boffetta
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): P. Boffetta
Writing, review, and/or revision of the manuscript: V. Galarraga, P. Boffetta
Study supervision: P. Boffetta

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 15, 2015; revised February 4, 2016; accepted March 1, 2016; published OnlineFirst March 28, 2016.

References