



**HAL**  
open science

# One-carbon metabolism and smoking related cancers

Anouar Fanidi

► **To cite this version:**

Anouar Fanidi. One-carbon metabolism and smoking related cancers. Santé publique et épidémiologie. Université de Lyon, 2016. English. NNT : 2016LYSE1002 . tel-01383446

**HAL Id: tel-01383446**

**<https://theses.hal.science/tel-01383446>**

Submitted on 18 Oct 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Année 2015

**THESE DE L'UNIVERSITE DE LYON**

**UNIVERSITE CLAUDE BERNARD – LYON 1**

Ecole Doctorale Interdisciplinaire Science et Santé (EDISS)

Pour l'obtention du

**DIPLOME DE DOCTORAT**

Mention « Epidémiologie, Santé publique, Recherche sur les services de santé »

**Soutenue publiquement le 7 Janvier 2016**

par

**Mr Anouar Fanidi**

***Le métabolisme monocarboné et les cancers liés au tabac***

***One-carbon metabolism and smoking-related cancers***

Thèse dirigée par Paul Brennan and Mattias Johansson

Jury composé de:

Pr	Mathoulin-Pelissier	Simone	Rapporteur
MD	Bethany	Van Guelpen	Examineur
Dr	Cécile	Delcourt	Rapporteur
Pr	Philippe	Vanhems	Examineur
Dr	Paul	Brennan	Directeur de thèse
Dr	Mattias	Johansson	Co-Directeur de thèse

# Résumé en français

Les vitamines B et les facteurs liés au métabolisme monocarboné (C1) aident à maintenir la synthèse de l'ADN, régulent l'expression des gènes, et peuvent affecter le risque de cancer. L'objectif général de cette thèse est d'étudier l'importance des biomarqueurs du C1 dans l'étiologie de trois cancers distincts qui diffèrent dans leur force d'association avec le tabagisme. Les articles inclus dans cette thèse ont été conduits au sein de deux études prospectives : l'étude prospective européenne sur la nutrition et le cancer (EPIC) et le consortium de cohortes du cancer du poumon (LC3).

Dans l'article 1, nous avons étudié la relation entre les biomarqueurs du C1 et l'incidence ainsi que le pronostic du cancer de la sphère oto-rhino-laryngée (ORL) et de l'œsophage au sein de l'étude EPIC. Nous avons observé que les sujets ayant des concentrations élevées d'homocystéine avaient un risque accru de développer un cancer de la sphère ORL.

Dans l'article 2, nous avons examiné la relation entre les biomarqueurs du C1 et l'incidence ainsi que le pronostic du carcinome à cellules rénales (CCR) au sein de l'étude EPIC. Nous avons constaté que les participants ayant des concentrations de vitamine B6 élevées avaient une diminution du risque de CCR avec un effet dose-réponse ainsi qu'une amélioration de la survie post-diagnostic.

Dans l'article 3, nous avons étudié si les biomarqueurs du C1 sont associés au risque de cancer du poumon au sein de l'étude LC3. Dans l'ensemble, nous avons mis en exergue une faible association inverse, sans tendance claire, entre les concentrations de vitamine B6 et de folate et le risque du cancer du poumon.

La principale conclusion de nos études est que les concentrations élevées de vitamine B6 sont associées à un risque plus faible de développer un CCR, et également à un meilleur pronostic chez les patients atteints de cette pathologie. Davantage d'études sont nécessaires afin d'évaluer si la vitamine B6 exerce une influence causale sur l'étiologie et la mortalité du CCR, ou si d'autres facteurs métaboliques sont impliqués.

Mots clés: métabolisme monocarboné, cancer du poumon, cancer de la tête et du cou, carcinome à cellules rénales, EPIC, LC3, étude prospective.

# English summary

B-vitamins and factors related to one-carbon metabolism (OCM) pathway help to maintain DNA synthesis and regulate gene expression and may affect cancer risk. The overarching aim of this thesis is to investigate the importance of OCM biomarkers in the etiology of three distinct cancer sites that differed in their strength of association with smoking. Papers included in this thesis were conducted within two prospective studies, the European Prospective Investigation into nutrition and Cancer (EPIC) study and the Lung Cancer Cohort Consortium (LC3).

In paper 1, we investigated if OCM biomarkers are associated with incidence and survival of cancer of the head and neck and esophagus in the EPIC study. We observed that subjects with higher concentrations of homocysteine had increased risk of developing head and neck cancer.

In paper 2, we investigated if OCM biomarkers are associated with incidence and survival of renal cell carcinoma (RCC) in the EPIC study. We observed that study subjects with elevated vitamin B6 concentrations had lower risk of RCC in a dose-response fashion and improved survival following diagnosis.

In paper 3, we investigated whether OCM factors are associated with lung cancer risk in the LC3 study. Overall, we observed a weak inverse association, with no clear trend, between concentrations of vitamin B6 and folate and risk of lung cancer.

The most important conclusion is that elevated vitamin B6 concentrations are associated with lower risk of developing RCC, and also better prognosis among RCC cases. Further studies are warranted to evaluate if vitamin B6 exerts a causal influence on RCC etiology and mortality, or if other metabolic factors are involved.

Key words: One-carbon metabolism, lung cancer, head and neck cancer, renal cell carcinoma, EPIC, LC3, prospective study

# UNIVERSITE CLAUDE BERNARD - LYON 1

## **Président de l'Université**

Vice-président du Conseil d'Administration

Vice-président du Conseil des Etudes et de la Vie Universitaire

Vice-président du Conseil Scientifique

Directeur Général des Services

**M. François-Noël GILLY**

M. le Professeur Hamda BEN HADID

M. le Professeur Philippe LALLE

M. le Professeur Germain GILLET

M. Alain HELLEU

## ***COMPOSANTES SANTE***

Faculté de Médecine Lyon Est – Claude Bernard

Faculté de Médecine et de Maïeutique Lyon Sud – Charles  
Mérieux

Faculté d'Odontologie

Institut des Sciences Pharmaceutiques et Biologiques

Institut des Sciences et Techniques de la Réadaptation

Département de formation et Centre de Recherche en Biologie  
Humaine

Directeur : M. le Professeur J. ETIENNE

Directeur : Mme la Professeure C. BURILLON

Directeur : M. le Professeur D. BOURGEOIS

Directeur : Mme la Professeure C. VINCIGUERRA

Directeur : M. le Professeur Y. MATILLON

Directeur : Mme. la Professeure A-M. SCHOTT

## ***COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE***

Faculté des Sciences et Technologies

Département Biologie

Département Chimie Biochimie

Département GEP

Département Informatique

Département Mathématiques

Département Mécanique

Département Physique

UFR Sciences et Techniques des Activités Physiques et Sportives

Observatoire des Sciences de l'Univers de Lyon

Polytech Lyon

Ecole Supérieure de Chimie Physique Electronique

Institut Universitaire de Technologie de Lyon 1

Ecole Supérieure du Professorat et de l'Education

Institut de Science Financière et d'Assurances

Directeur : M. F. DE MARCHI

Directeur : M. le Professeur F. FLEURY

Directeur : Mme Caroline FELIX

Directeur : M. Hassan HAMMOURI

Directeur : M. le Professeur S. AKKOUCHE

Directeur : M. le Professeur Georges TOMANOV

Directeur : M. le Professeur H. BEN HADID

Directeur : M. Jean-Claude PLENET

Directeur : M. Y. VANPOULLE

Directeur : M. B. GUIDERDONI

Directeur : M. P. FOURNIER

Directeur : M. G. PIGNAULT

Directeur : M. le Professeur C. VITON

Directeur : M. le Professeur A. MOUGNIOTTE

Directeur : M. N. LEBOISNE

*To my parents,*

THIS THESIS HAS BEEN PREPARED AT

**International Agency for Research on Cancer (IARC)**

Section of Genetic Epidemiology

150 cours Albert Thomas

69372 Lyon Cedex 08

France

# Acknowledgement

*This work was possible thanks to many people. In particular, I want to acknowledge the following:*

**Paul Brennan** - my supervisor - to have given me the opportunity to do my PhD in such a privileged collaboration between France and the international. Thanks for your trust and to have introduced me to a large network of collaborators but also to have leaved me grow professionally in giving me autonomy and responsibility in project management.

**Mattias Johansson** - my co-supervisor - for welcoming me so warmly in the Genetic Epidemiology Group. Here, is all my gratitude for being particularly supportive and teaching me about methods in epidemiology. Thanks for your availability and bright advices.

**Devasena and Magdalena** - my two office mates or I should say "sisters" - for your smiles and teaching about principles in biology.

**David C Muller** – my dear friend and colleague – for your cancer epidemiological and statistical insights and mentoring through the peaks and pitfalls of epidemiology.

Charlotte Volatier- for helping me with all administrative issues. You really made my life easier many times.

**Helene Renard** – for helping me to solve informatics issues and improve my skills in data management.

**All my other colleagues** from the GEP group for your warm welcome, for inspiring and helping me to finish this thesis.

**All study participants** in EPIC and LC3 studies. Without your generous contribution to medical research, the studies underlying this thesis would never have been done.

**Doudou** – my sister in law – for your wonderful and delicious food and pastries. Nothing is better than a sweetness to forget the stress of the thesis.

**My brothers** for your love and for supporting me over the last years giving me energy and faith when I needed you the most.

**My parents** for your unconditional support and love. A particular thanks to my own flesh on blood, to this immeasurably affectionate look, my mother, a person who never doubted me and always prayed for me. This thesis is specifically dedicated to you.



*Ce travail a été possible grâce à de nombreuses personnes. En particulier, je tiens à remercier les suivantes:*

**Paul Brennan** - mon superviseur – qui m’a donné l’occasion de poursuivre mon doctorat dans un tel institut qu’est le Centre International de Recherche sur le Cancer. Merci pour ta confiance et pour m’avoir introduit dans un vaste réseau de collaborateurs. Merci également de m’avoir aidé à grandir professionnellement en me donnant de l’autonomie et de la responsabilité dans la gestion de projet.

**Mattias Johansson** - mon co-superviseur - de m’avoir accueilli si chaleureusement dans le Groupe d’Épidémiologie Génétique (GEP). Trouve ici toute ma reconnaissance pour m’avoir soutenu et pour tes enseignements riches concernant les méthodes en épidémiologie. Un grand merci également pour ta disponibilité et tes conseils brillants.

Mes « sœurs de thèse » - **Devasena et Magdalena** - pour vos étincelants sourires et votre aide concernant les principes de la biologie.

**David C Muller** - mon cher ami et collègue - pour ton enseignement des statistiques et de l’épidémiologie des cancers. Merci de m’avoir soutenu et supervisé tout au long des « chemins ardu et piégés » de l’épidémiologie.

**Charlotte Volatier**- Un grand merci pour ton aide concernant toutes les questions administratives. Tu as vraiment facilité ma vie à plusieurs reprises.

**Hélène Renard** – pour ton aide précieuse quant à résoudre les problèmes informatiques. Merci aussi de m’avoir aidé à améliorer mes compétences en data management.

**Tous mes autres collègues du groupe GEP** pour votre accueil si chaleureux et pour ceux qui ont contribué à finaliser cette thèse.

**Tous les participants des études EPIC et LC3.** Sans votre généreuse contribution à la recherche médicale, les études sous-jacentes de cette thèse n’auraient pas vues le jour.

**Doudou** - ma belle-sœur - pour tes délicieuses et succulentes pâtisseries. Rien n’est meilleur qu’une petite douceur pour oublier le stress de la thèse.

**Mes frères** pour votre amour et pour m’avoir soutenu lors de ces dernières années en me donnant de l’énergie et de la confiance quand j’en avais le plus besoin.

**Mes parents** pour votre soutien inconditionnel et votre amour. Un remerciement particulier à la chair de ma chair, à ce regard incommensurablement affectueux - ma mère - une personne qui n’a jamais douté de moi et qui a toujours prié pour moi. Cette thèse est spécifiquement dédiée à toi.

# Résumé substantiel en français

Les cancers liés au tabagisme sont définis comme les cancers causés par des expositions au tabac. À ce jour, plus de 3000 constituants ont été isolés à partir du tabac. Les avancées technologiques ont permis l'identification d'environ 70 substances cancérigènes dans la fumée du tabac, ainsi qu'une substance puissamment addictive, la nicotine. La fumée des produits du tabac, et en particulier des cigarettes est le facteur de risque le plus important pour le cancer. Il a été montré que cette exposition est clairement associée à un tiers de tous les cancers à l'échelle mondiale incluant les cancers du poumon (70-80% de tous les cas), de la tête et du cou (60-70%), du foie (50%) et du rein (30%). Ces connaissances ont été traduites par l'évitement du tabac et la mise en place de programmes de sevrage, ce qui a conduit à des réductions d'incidence de cancers liés au tabagisme (notamment le cancer du poumon) chez les hommes dans la plupart des pays européens. Même chez les sujets qui ont cessé de fumer, le risque à vie de cancer lié au tabagisme reste élevé. De plus, dans certaines populations où les campagnes anti-tabac ont été couronnées de succès, comme aux États-Unis, près de 50% des cas de cancer du poumon surviennent chez les non et anciens fumeurs. Il paraît essentiel d'identifier des facteurs de risque afin d'en dégager des informations supplémentaires inter et intra groupes de fumeurs. Les conclusions mises en évidence revêtiraient un intérêt majeur pour la santé publique.

L'alimentation et la nutrition sont supposées jouer un rôle fondamental dans le développement des cancers les plus courants. Il a été estimé qu'une mauvaise alimentation, en particulier une faible consommation de fruits et légumes, contribuerait à environ un tiers des cancers évitables. Les fruits et légumes sont des sources de vitamines B et d'autres micronutriments pour le métabolisme monocarboné (C1). Des carences et/ou déséquilibres de micronutriments dans la voie du C1 peuvent affecter le risque de cancer par le biais de multiples mécanismes tels que les méthylations anormales d'ADN ou des altérations dans les processus de réparation et de synthèse de l'ADN. Malgré un certain nombre d'études épidémiologiques suggérant un rôle possible du folate (vitamine B9) dans la prévention du cancer, en particulier le cancer colorectal, les preuves suggérant un lien de causalité entre le C1 et la carcinogenèse sont limitées. De fortes associations inverses entre des

concentrations pré-diagnostiques élevées de biomarqueurs du C1 (notamment la vitamine B6, le folate, et la méthionine) et le risque du cancer du poumon ont été récemment signalées dans une grande étude cas-témoins nichée au sein de l'étude EPIC. L'étude a notamment révélé des associations fortes et cohérentes parmi les non et anciens fumeurs, suggérant que les résultats n'étaient probablement pas dus au facteur de confusion qu'est le tabagisme.

L'objectif général de cette thèse est d'étudier l'importance des biomarqueurs du C1 dans l'étiologie de trois cancer distincts (sphère ORL, rein et poumon) qui diffèrent dans leur force d'association avec le tabagisme.

Dans l'article 1, nous avons étudié la relation entre les concentrations circulantes des facteurs du C1 et l'incidence ainsi que le pronostic du cancer de la sphère ORL et de l'œsophage. Un échantillon de 516 cas et 516 témoins appariés individuellement sur le pays, le sexe, la date de naissance, le pays et le prélèvement de sang a été recruté à partir de l'étude EPIC. Les concentrations plasmatiques de vitamines B2, B6, B9 (folate), B12, d'homocystéine et de méthionine ont été mesurées à partir d'échantillons pré-diagnostiques sanguins. Nous avons remarqué que les sujets ayant des concentrations élevées d'homocystéine avaient un risque accru de développer un cancer de la tête et du cou. L'association positive de l'homocystéine était présente à la fois chez les non-fumeurs et les non-consommateurs d'alcool. Les concentrations circulantes des autres biomarqueurs ne présentaient aucun lien évident avec le risque ou la survie.

Dans l'article 2, nous voulions évaluer l'ampleur avec laquelle l'homocystéine ou d'autres biomarqueurs inclus dans la voie du C1 pouvait être liée à l'incidence et au pronostic du carcinome à cellules rénales (CCR). Une raison supplémentaire d'étudier l'association entre les biomarqueurs du C1 et les CCR est que ce type de cancer présente une force d'association avec le tabagisme moindre comparé au cancer de la tête et du cou. Les concentrations plasmatiques de vitamines B2, B6, B9, B12, d'homocystéine et de méthionine ont été mesurées à partir d'échantillons pré-diagnostiques sanguins auprès de tous les participants de cette étude. Dans un premier temps nous avons recruté 556 paires cas-témoins au sein de la cohorte EPIC. Globalement, nous avons observé que les sujets d'EPIC ayant des concentrations de vitamine B6 élevées avaient moins de risque de CCR, de

manière dose-réponse, résultant à une diminution du risque approximative de 60% lorsque les quartiles supérieurs et inférieurs de la population étaient comparés. Nous avons également remarqué une claire amélioration de la survie auprès des cas atteints de CCR ayant des concentrations de vitamine B6 plus élevées. Afin de s'assurer que la relation entre la vitamine B6 et le carcinome à cellules rénales n'était pas restreinte à l'étude EPIC, une étude de réplique a été menée dans une cohorte distincte d'Australie, la Melbourne Collaborative Cohort Study (MCCS) incluant 144 paires cas-témoins. Nous avons obtenu des résultats significativement similaires à ceux observés dans la cohorte EPIC concernant les risques de CCR ainsi que la mortalité par cancer.

Dans l'article 3, nous voulions valider certaines conclusions antérieurement citées et observées au sein de la cohorte EPIC, impliquant de fortes associations inverses entre le folate, la vitamine B6 et la méthionine et le risque de cancer du poumon. Nous avons donc étudié l'association de ces trois biomarqueurs avec le risque de cancer du poumon au sein d'un vaste consortium, le consortium de cohortes du cancer du poumon (LC3). Ce dernier rassemble plusieurs populations d'Amérique, d'Europe, d'Asie et d'Australie. Au total, vingt études cas-témoins nichées ont contribué à cette étude avec des échantillons sanguins pré-diagnostiques de vitamines B6, B9 et de méthionine mesurées auprès des 5 364 paires de cas-témoins. Les critères d'appariement étaient la cohorte d'appartenance, le sexe, la date du prélèvement de sang, la date de naissance et le statut tabagique. Dans l'ensemble, nous avons observé de faibles associations inverses, sans tendance claire, entre les concentrations de vitamine B6, de folate et le risque de cancer du poumon, associations principalement présentes chez les fumeurs anciens et quotidiens. Réciproquement, les fumeurs anciens et quotidiens ayant un déficit en concentrations circulantes de vitamine B6 et de folate comparés à ceux ayant des concentrations normales ont montré une faible augmentation du risque du cancer du poumon.

En résumé, nos études suggèrent que les concentrations circulantes d'homocystéine affectent le risque de développer un cancer de la tête et du cou. Ces résultats nouveaux doivent être répliqués dans d'autres études prospectives de grandes tailles ainsi que dans différentes populations ethniques.

Les résultats de l'étude LC3 ne soutiennent pas l'association inverse entre les concentrations circulantes élevées de vitamine B6 et la diminution du risque de cancer du poumon, comme

précédemment indiqué dans l'étude EPIC du poumon. Cependant, nous avons remarqué de fortes associations entre le déficit en vitamine B6 et le risque du cancer du poumon suivant les trois ans après le diagnostic chez des fumeurs quotidiens. Cette observation soulève l'hypothèse que la vitamine B6 circulante pourrait être un potentiel prédicteur de risque chez les fumeurs quotidiens ayant été diagnostiqués proches de l'échantillon sanguin.

La principale conclusion de nos recherches est que les participants de l'étude EPIC ayant des concentrations circulantes de vitamine B6 élevées ont un moindre risque de développer un carcinome à cellules rénales tout en ayant un meilleur pronostic une fois cette maladie diagnostiquée. Cependant, des études complémentaires sont nécessaires pour évaluer si la vitamine B6 exerce une influence causale sur l'étiologie et la mortalité du CCR, ou si d'autres facteurs métaboliques sont impliqués.

# LIST OF TABLES

<b>Table 1.</b> Associations between OCM factors (vitamin B2, B6, B12, folate, methionine and homocysteine) and lung cancer risk .....	<b>47</b>
<b>Table 2.</b> Associations between OCM factors (vitamin B2, B6, B12, folate, methionine and homocysteine) and the head and neck cancer risk .....	<b>50</b>
<b>Table 3.</b> Associations between OCM biomarkers factors (vitamin B2, B6, folate and methionine) and esophageal squamous cell carcinoma risk.....	<b>52</b>
<b>Table 4.</b> Associations between OCM factors (vitamin B2, B6, B12, folate, methionine and homocysteine) and kidney cancer risk .....	<b>54</b>
<b>Table 5.</b> Characteristics of participating cohorts and lung cancer cases with serum/plasma available .....	<b>70</b>

# LIST OF FIGURES

<b>Figure 1.</b> Age-standardized incidence and mortality rates (world) per 100,000 for lung cancer in 2012. Source: Globocan 2012 <sup>15</sup> .	<b>25</b>
<b>Figure 2.</b> Observed 5-years survival based in patients diagnosed with NSCLC in 2004-2009 on the SEER 9 areas (San Francisco, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, and Atlanta), adapted from the SEER database <sup>9</sup> .	<b>27</b>
<b>Figure 3.</b> 5-year survival by stage for people diagnosed with lung cancer in 2004-2009 on the SEER 9 areas (San Francisco, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, and Atlanta), adapted from the SEER database <sup>8</sup> .	<b>27</b>
<b>Figure 4.</b> Schematic of various head and neck malignancies. Source from Terese Winslow LLC. .	<b>29</b>
<b>Figure 5.</b> Age-standardized incidence and mortality rates (world) per 100,000 for head and neck cancer (lip, oral cavity, larynx, nasopharynx and other pharynx) in 2012. Source: Globocan 2012 <sup>15</sup> .	<b>30</b>
<b>Figure 6.</b> 5-year relative survival of HNC subsites in US, adapted from the SEER database <sup>49</sup> .	<b>31</b>
<b>Figure 7.</b> Age-standardised incidence rates per 100, 000 of oesophageal cancers by histological subtype, region and sex. AC, adenocarcinoma; SCC, squamous cell carcinoma. Source from Arnold M et al <sup>71</sup> .	<b>34</b>
<b>Figure 8.</b> Age-standardized incidence and mortality rates (world) per 100,000 for kidney cancer in 2012. Source: Globocan 2012 <sup>15</sup> .	<b>37</b>
<b>Figure 9.</b> Survival estimates in 4,063 patients with renal tumours according to histologic subtype and TNM stage. (A) TNM stage I to II. (B) TNM stage III to IV. Red dashed lines correspond to five-year survival. Source from Patard JJ et al. <sup>105</sup> .	<b>38</b>
<b>Figure 10.</b> Overview of the OCM pathway. The OCM consists of the folate cycle (green) and the methionine cycle (gray). Cofactors are in purple and enzymes are in orange. SHMT: serine hydroxymethyltransferase, TS: thymidylate synthase, MTHFR: methylenetetrahydrofolate reductase, MAT: methionine adenosyltransferase, CBS: cystathionine-synthase, CGL: cystathionine $\gamma$ -lyase, MS: methionine synthase, THF: tetrahydrofolate, me-THF: 5,10-methylenetetrahydrofolate, mTHF: 5-methyltetrahydrofolate, F-THF:10-formyltetrahydrofolate, dTMP: deoxythymidine monophosphate, dUMP: deoxyuridine monophosphate, SAH: S-adenosylhomocysteine, SAM: S-adenosylmethionine, B2: riboflavin, B6: pyridoxal 5'-phosphate, B12: cobalamin, adapted from Locasale et al., 2013 <sup>133</sup> .	<b>41</b>
<b>Figure 11.</b> Representation of risk sets in nested case-control designs. .	<b>64</b>
<b>Figure 12.</b> EPIC study collaborating centres. .	<b>67</b>

# APPENDICES

<b>Appendix 1-</b> Supplementary methods of LC3 cohorts .....	<b>162</b>
<b>Appendix 2-</b> Data dictionary of LC3 variables .....	<b>176</b>
<b>Appendix 3-</b> Supplementary table 1 for Article 1 .....	<b>193</b>
<b>Appendix 4-</b> Supplementary table 2 for Article 1 .....	<b>194</b>
<b>Appendix 5-</b> Supplementary table 3 for Article 1 .....	<b>195</b>
<b>Appendix 6-</b> Supplementary table 4 for Article 1 .....	<b>196</b>
<b>Appendix 7-</b> Supplementary table 5 for Article 1 .....	<b>197</b>
<b>Appendix 8-</b> Supplementary figure 1 for Article 1 .....	<b>198</b>
<b>Appendix 9-</b> Supplementary figure 2 for Article 1 .....	<b>199</b>
<b>Appendix 10-</b> Supplementary figure 3 for Article 1 .....	<b>200</b>
<b>Appendix 11-</b> Supplementary figure 4 for Article 1 .....	<b>201</b>
<b>Appendix 12-</b> Supplementary figure 5 for Article 1 .....	<b>202</b>
<b>Appendix 13-</b> Supplementary figure 6 for Article 1 .....	<b>203</b>
<b>Appendix 14-</b> Supplementary methods for Article 1 .....	<b>204</b>
<b>Appendix 15-</b> Supplementary table 1 for Article 2 .....	<b>206</b>
<b>Appendix 16-</b> Supplementary table 2 for Article 2 .....	<b>207</b>
<b>Appendix 17-</b> Supplementary table 3 for Article 2 .....	<b>208</b>
<b>Appendix 18-</b> Supplementary table 4 for Article 2 .....	<b>209</b>
<b>Appendix 19-</b> Supplementary table 5 for Article 2 .....	<b>210</b>
<b>Appendix 20-</b> Supplementary figure 1 for Article 3 .....	<b>212</b>
<b>Appendix 21-</b> Supplementary figure 2 for Article 3 .....	<b>213</b>
<b>Appendix 22-</b> Supplementary figure 3 for Article 3 .....	<b>214</b>
<b>Appendix 23-</b> Supplementary figure 4 for Article 3 .....	<b>215</b>
<b>Appendix 24-</b> Supplementary figure 5 for Article 3 .....	<b>216</b>
<b>Appendix 25-</b> Supplementary figure 6 for Article 3 .....	<b>217</b>
<b>Appendix 26-</b> Supplementary figure 7 for Article 3 .....	<b>218</b>
<b>Appendix 27-</b> Supplementary figure 8 for Article 3 .....	<b>219</b>
<b>Appendix 28-</b> Supplementary figure 9 for Article 3 .....	<b>220</b>
<b>Appendix 29-</b> Supplementary figure 10 for Article 3 .....	<b>221</b>
<b>Appendix 30-</b> Supplementary figure 11 for Article 3 .....	<b>222</b>
<b>Appendix 31-</b> Supplementary figure 12 for Article 3 .....	<b>223</b>
<b>Appendix 32-</b> Supplementary figure 13 for Article 3 .....	<b>224</b>
<b>Appendix 33-</b> Supplementary figure 14 for Article 3 .....	<b>225</b>
<b>Appendix 34-</b> Supplementary table 1 for Article 3 .....	<b>226</b>
<b>Appendix 35-</b> Supplementary table 2 for Article 3 .....	<b>228</b>
<b>Appendix 36-</b> Supplementary table 3 for Article 3 .....	<b>230</b>
<b>Appendix 37-</b> Supplementary table 4 for Article 3 .....	<b>232</b>



# ABBREVIATIONS

OCM	one-carbon metabolism
EPIC	European Prospective Investigation into Cancer and Nutrition
HNC	head and neck cancer
RCC	renal cell carcinoma
LC3	Lung Cancer Cohort Consortium
ASR	age-standardised rate
SCLS	small cell lung carcinoma
SCC	squamous cell carcinoma
NSCLC	non-small cell lung carcinoma
TNM	Classification of Malignant Tumours
IARC	International Agency for Research on Cancer
EPA	Environmental Protection Agency
RCT	randomized controlled trial
ATBC	Alpha-Tocopherol, Beta-Carotene, Prevention Cancer
WCRF/AICR	World Cancer Research Fund/American Institute for Cancer Research
HNSCC	head and neck cancer are squamous cell carcinoma
NLCS	Netherlands Cohort Study
HPV	human papilloma virus
WCRF	World Cancer Research Fund
SEER	Surveillance, Epidemiology and End Results Program
THF	tetrahydrofolate
TS	thymidylate synthase
MTHFR	methylenetetrahydrofolate réductase
MAT	methionine adenosyltransferase
CBS	cystathionine-synthase
CGL	cystathionine $\gamma$ -lyase

MS	methionine synthase
me-THF	5,10-methylenetetrahydrofolate
mTHF	5-methyltetrahydrofolate
F-THF	10-formyltetrahydrofolate
dTMP	deoxythymidine monophosphate
dUMP	deoxyuridine monophosphate
SAH	S-adenosylhomocysteine
SAM	S-adenosylmethionine
SHMT	serine hydroxymethyl transferase
SHMT	serine hydroxymethyltransferase
CBS	cystathionine $\beta$ -synthase
CGL	cystathionine $\gamma$ -lyase
CpG	cytosine-phosphate-Guanine
CRC	colorectal cancer
INHANCE	International Head and Neck Cancer Epidemiology
OR	odds ratio
RR	relative risk
ICD	International Classification of Diseases
CVs	coefficient variations
ICC	intra-class correlation

# TABLE OF CONTENTS

<b>LIST OF ORIGINAL PUBLICATIONS .....</b>	<b>21</b>
<b>CHAPTER I: BACKGROUND AND INTRODUCTION .....</b>	<b>22</b>
<b>Overview and aim .....</b>	<b>23</b>
<b>Smoking related-cancers - epidemiologic evidence .....</b>	<b>25</b>
Lung cancer .....	25
Incidence and mortality .....	25
Risk factors of lung cancer .....	27
Head and neck cancer .....	29
Incidence and mortality .....	29
Risk factors of head and neck cancer.....	31
Esophageal cancer .....	33
Incidence and mortality .....	33
Risk factors of esophagus squamous cell carcinoma .....	34
Kidney cancer .....	36
Incidence and mortality .....	36
Risk factors of kidney cancer .....	38
<b>One-carbon metabolism: a candidate pathway.....</b>	<b>40</b>
Overview of the one-carbon metabolism (OCM) .....	40
OCM and carcinogenesis.....	42
DNA repair and DNA synthesis.....	42
DNA methylation.....	43
DNA methylation and cancer .....	44
OCM and methylation.....	45
<b>Nutrients of the candidate pathway and smoking-related cancer .....</b>	<b>46</b>
Overview .....	46
OCM analytes and lung cancer .....	46
OCM factors and head and neck cancer .....	49
OCM factors and esophageal squamous cell carcinoma .....	51
OCM factors and kidney cancer .....	53
<b>Study designs in analytical epidemiology .....</b>	<b>55</b>
Experimental studies.....	55

Cohort studies.....	56
Case-control studies.....	59
Nested case-control studies.....	63
<b>CHAPTER II: MATERIAL AND METHODS.....</b>	<b>66</b>
<b>Data sources .....</b>	<b>67</b>
The European Prospective Investigation into nutrition and Cancer (EPIC) .....	67
The Lung Cancer Cohort Consortium (LC3).....	69
<b>Biochemical analyses .....</b>	<b>71</b>
<b>Statistical methods.....</b>	<b>73</b>
<b>CHAPTER III: ONE-CARBON METABOLISM AND HEAD AND NECK CANCER .....</b>	<b>76</b>
<b>Objective .....</b>	<b>77</b>
<b>Selection of cases and controls .....</b>	<b>77</b>
<b>Biochemical analyses .....</b>	<b>78</b>
<b>Statistical analyses .....</b>	<b>78</b>
<b>Article 1 .....</b>	<b>79</b>
<b>CHAPTER IV: ONE-CARBON METABOLISM AND RENAL CELL CARCINOMA.....</b>	<b>93</b>
<b>Objective .....</b>	<b>94</b>
<b>Selection of cases and controls .....</b>	<b>94</b>
<b>Replication study .....</b>	<b>95</b>
<b>Biochemical analyses .....</b>	<b>95</b>
<b>Statistical analyses .....</b>	<b>95</b>
<b>Article 2 .....</b>	<b>96</b>
<b>CHAPTER V: B-VITAMINS AND LUNG CANCER IN THE LC3 .....</b>	<b>108</b>
<b>Objective .....</b>	<b>109</b>
<b>Selection of cases and controls .....</b>	<b>109</b>
<b>Data harmonization .....</b>	<b>109</b>
<b>Biochemical analyses .....</b>	<b>110</b>
<b>Statistical analyses .....</b>	<b>110</b>
<b>Article 3 .....</b>	<b>111</b>
<b>CHAPTER VI: DISCUSSION .....</b>	<b>134</b>
<b>The one-carbon metabolism and smoking-related cancers.....</b>	<b>135</b>
Recall on objectives .....	135

Summary of the results.....	135
Observations across the studies .....	138
<b>Main strengths and limitations .....</b>	<b>142</b>
<b>Future directions.....</b>	<b>144</b>
<b>REFERENCES .....</b>	<b>147</b>
<b>APPENDICES .....</b>	<b>162</b>

## LIST OF ORIGINAL PUBLICATIONS

**Fanidi A**, Relton C, Ueland P.M *et al.* A prospective study of one-carbon metabolism biomarkers and cancer of the head and neck and esophagus. *Int J Cancer*. 2015 Feb 15;136(4):915-27

Johansson M\*, **Fanidi A\***, Muller DC\* *et al.* Circulating biomarkers of one-carbon metabolism in relation to renal cell carcinoma incidence and survival. *J Natl Cancer Inst*. 2014 Nov 5;106(12):1-11. (\*equally contributed).

**Anouar Fanidi\***, David Muller\*, Jian-Min Yuan\*, Victoria L. Stevens\*, Stephanie J. Weinstein\*, Demetrius Albanes\*, Ross Prentice\*, Cynthia A. Thomsen, Mary Pettinger, Qiuyin Cai, William J. Blot, Jie Wu, Alan A. Arslan, Anne Zeleniuch-Jacquotte,, Marjorie L. McCullough, Loic Le Marchand, Lynne R. Wilkens, Christopher A. Haiman, Xuehong Zhang, Jiali Han, Meir J. Stampfer, Stephanie A. Smith-Warner, Edward Giovannucci, Graham G. Giles, Allison M Hodge, Gianluca Severi, Mikael Johansson, Kjell Grankvist, Arnulf Langhammer, Steinar Krokstad, Marit Næss, Renwei Wang, Yu-Tang Gao, Lesley M. Butler, Woon-Puay Koh, Xiao-Ou Shu, Yong-Bing Xiang, Honglan Li, Wei Zheng, Qing Lan, Kala Visvanathan, Judith Hoffman Bolton, Per Magne Ueland, Øivind Midttun, Arve Ulvik, Neil Caporaso, Mark Purdue, Regina Zeigler, Neal D. Freedman, Julie E. Buring, I-Min Lee, Howard D. Sesso, J. Michael Gaziano, Jonas Manjer, Ulrika Ericson, Caroline Relton, Paul Brennan\*\*, Mattias Johansson\*\*. Circulating Folate, Vitamin B6 and Methionine in relation to Lung Cancer Risk in the Lung Cancer Cohort Consortium (LC3).

# CHAPTER I:

## *BACKGROUND AND INTRODUCTION*

## Overview and aim

Smoking-related cancers are defined as cancers caused by tobacco exposures. Over 3000 constituents have been isolated from tobacco and advancing technologies have permitted the identification around 70 carcinogens in tobacco smoke to date, as well as a powerfully addictive substance, nicotine<sup>1</sup>. Approximately 1.1 billion people worldwide smoke with important sex-differences. In many countries, smoking rates are increasing in young women, with potentially grave consequences for cancer<sup>2</sup>. Smoking of tobacco products, and in particular cigarettes is the most important risk factor for cancer and has been shown to be clearly associated with about one in three of all cancers globally<sup>3</sup>, including lung (70-80% of all cases)<sup>4</sup>, head and neck (60-70%,)<sup>5</sup>, liver (50%)<sup>6</sup> and kidney (30%)<sup>7</sup>. This knowledge has been translated into tobacco avoidance and cessation programs, which has led to reductions in smoking-related cancer (most notably lung cancer) incidence among men in most European countries<sup>8</sup>. While it has been observed that in populations where the large proportion of the population are no longer current smokers, the majority of smoking-related cancer cases continue to occur among former smokers. In addition, survival from smoking-related cancers is poor, with 5-year mortality in the region of approximately 85% for lung cancer and 50% for other cancers sites such as head and neck cancer<sup>9</sup>. Identifying modifiable additional risk factors within and across smoking groups remains of great clinical and public health interest.

Diet and nutrition are thought to play a fundamental role in the development of most common cancers<sup>10</sup>. It has been estimated that poor diet, in particular low intake of fruit and vegetables contribute to about one-third of preventable cancers<sup>11</sup>. Fruits and vegetables are sources of B-vitamins and other metabolites of the one-carbon metabolism (OCM) pathway. Deficiencies and/or imbalances of the micronutrients in the OCM pathway may impact cancer risk through multiple mechanisms such as aberrant DNA methylation patterns or DNA damage through an oxidative imbalance. Despite a number of epidemiological studies suggesting a possible role of folate (vitamin B9) in cancer prevention, particularly for colorectal cancer, evidence supporting a causal relationship between OCM and carcinogenesis is lacking. Strong inverse associations of elevated pre-diagnostic circulating concentrations of OCM biomarkers (in particular vitamin B6, folate, and also methionine) with lung cancer risk have been recently reported in a large nested case-control study within the European Prospective Investigation into Cancer and Nutrition (EPIC)<sup>12</sup>. The study found



strong and consistent associations for never- and former smokers, suggesting that the result was unlikely to be due to confounding by smoking.

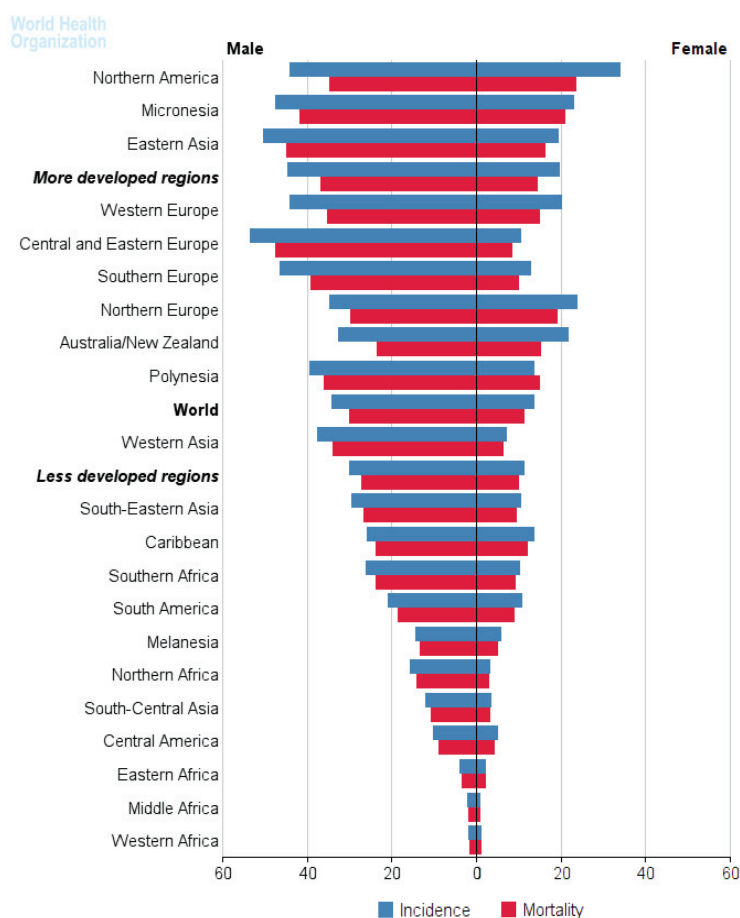
The overarching aim of this thesis is to investigate the importance of OCM biomarkers in the etiology of three distinct cancer sites (head and neck, kidney and lung) that differed in their strength of association with smoking. The head and neck cancer (HNC) and the renal cell carcinoma (RCC) studies investigated in EPIC cohort whether pre-diagnostic circulating markers of the one-carbon pathway including vitamin B2, vitamin B6, folate, vitamin B12, homocysteine and methionine are related to incidence and survival of cancer of the head and neck and squamous cell esophagus (Chapter III), and RCC (Chapter IV), in a similar or different way as for lung cancer. The Lung Cancer Cohort Consortium (LC3) study (Chapter V) evaluated whether pre-diagnostic circulating concentrations of vitamin B6, folate and methionine are associated with lung cancer risk in a worldwide consortium of prospective cohort studies, in order to assess the generalisability of the recently published findings<sup>13</sup>.

## Smoking related-cancers - epidemiologic evidence

### Lung cancer

#### *Incidence and mortality*

Lung cancer has been reported as the most common cancer globally, with approximately 1,800,000 new cases diagnosed annually<sup>14</sup>. In 2012, lung cancer accounted for an estimated 240,000 new cases in the United States (US), representing 15% of lung cancers worldwide<sup>14</sup>. However, incidence rates of lung cancer display remarkable variation across the world (see [Figure 1](#))<sup>15</sup>.



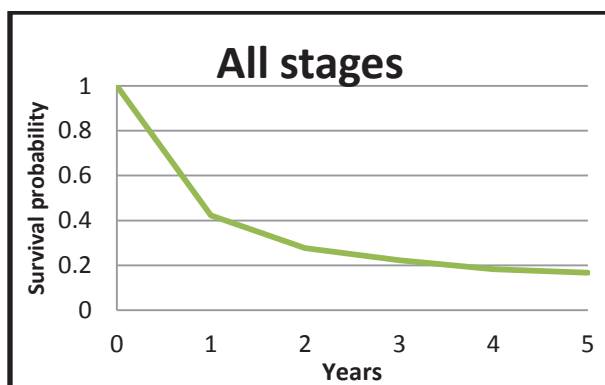
**Figure 1.** Age-standardized incidence and mortality rates (world) per 100,000 for lung cancer in 2012. Source: *Globocan 2012*<sup>15</sup>.

Among men, the highest estimated age-standardised incidence rates (ASR) are observed in Central and Eastern Europe (53.5 per 100,000), Eastern Asia (50.4 per 100,000) and North America (48.5 per 100,000) and the lowest ASR are located in sub-Saharan Africa (4.8 per 100,000)<sup>15</sup>. Similar to men, the highest incidence rates among women are also observed in North America (33.8 per 100,000) and Eastern Asia (19.2 per 100,000) while the lowest ASR

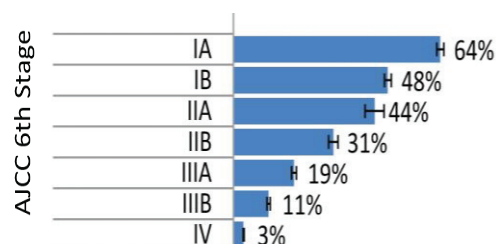
are also located in sub-Saharan Africa (2.5 per 100,000). When comparing incidence rates by sex, rates are generally lower in women than in men, mainly reflecting different historical exposure to tobacco smoking where women started smoking in large numbers several decades later than men<sup>16</sup>. However, recent trends show that lung cancer among women is increasing considerably in many countries including European countries<sup>17</sup>.

Lung cancer has an enormous impact on worldwide mortality, estimated to be responsible for nearly 1.6 million deaths worldwide in 2012<sup>2</sup>. Geographical patterns of mortality are quite similar to those of incidence for both sexes (see [Figure 1](#)) due to the rapid course of the disease after diagnosis, then reflecting the poor 5-year survival. Lung cancer mortality among men is decreasing in most Western countries, including many European countries, where the tobacco epidemic diminished after the middle of the last century<sup>17-19</sup>. By contrast, it has risen sharply in many low- and middle-income countries who were at an earlier stage in the smoking epidemic, such as China and Indonesia<sup>2</sup>.

Lung cancer five-year relative survival rate for all stages combined is about 15% in the United States and Europe<sup>9 20 21</sup>. Survival is strongly influenced by histology and stage at diagnosis. The main histological types of lung cancer encompass small cell lung carcinoma (SCLC), squamous cell carcinoma (SCC), adenocarcinoma and large cell carcinoma (the latter three types being commonly referred as non-small cell lung carcinoma, or NSCLC). SCLC, representing less than 20% of the overall lung cancer burden, is usually classified as limited or extensive and is generally responsive to therapy<sup>22</sup>. In contrast, NSCLC representing over 80% of lung cancers are categorized using the TNM Classification of Malignant Tumours into various stages<sup>23</sup>. Recent data from US (see [Figure 2](#)) shows that overall 5-year survival of lung cancer is over 20%. When stratifying by stage (see [Figure 3](#)), it has been observed that cases diagnosed at early stages have a 5-year survival ranging from 30%-60% (stage I and II). In contrast, cases diagnosed at a later stage (i.e. IV) have very poor prognosis with 5-year survival of below 5%.



**Figure 2.** Observed 5-years survival based in patients diagnosed with NSCLC in 2004-2009 on the SEER 9 areas (San Francisco, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, and Atlanta), adapted from the SEER database<sup>9</sup>.



**Figure 3.** 5-year survival by stage for people diagnosed with lung cancer in 2004-2009 on the SEER 9 areas (San Francisco, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, and Atlanta), adapted from the SEER database<sup>8</sup>.

### **Risk factors of lung cancer**

Epidemiological evidence of the association between tobacco smoking and cancer began to emerge in the 1920's<sup>24</sup> and by the 1950's the causal relationship with lung cancer was established by Doll and Hill<sup>25 26</sup>. They reported that smokers who smoked 25 or more cigarettes per day had 50 times increased risk of lung cancer compared to non-smokers. This work was followed by several epidemiological studies, and by the mid-1960s, the causal association between lung cancer and smoking was recognized by public health and regulatory authorities. In 1986, the International Agency for Research on Cancer (IARC) monograph also reported the causality of tobacco smoking with lung cancer<sup>27</sup>. The excess risk of lung cancer among smokers relative to non-smokers depends on several aspects including intensity of smoking, duration of smoking, type of smoking product and age at start<sup>28 29</sup>.

While tobacco exposure is well known to be strongly associated with lung cancer in ever smokers, there is still limited evidence of the causes of lung cancer in never smokers. Secondhand smoke has been established as a major risk factor among never smokers<sup>30</sup>. Also, other factors including indoor pollutants (coal, wood fire indoors)<sup>31 32</sup> and outdoor air pollution<sup>33</sup> may affect lung cancer development.

Additional chemical substances have been linked to lung cancer, such as radon. According to the US Environmental Protection Agency (EPA), this radioactive gas causes about 20,000 cases of lung cancer each year<sup>34 35</sup>. There is also convincing evidence from several ecological, case-control and cohort studies showing that arsenic in drinking water has a causal relationship with lung cancer<sup>36 37</sup>. Other substances found to increase lung cancer risk include asbestos and silica<sup>37</sup>.

Recent conclusion from the IARC monograph stated that there was possible but not established evidence for carcinogenicity of alcohol consumption with lung cancer<sup>38</sup>. However, it is difficult to disentangle the role of alcohol in lung cancer development given the strong correlation between alcohol drinking and tobacco smoking. Heavy drinkers tend to be heavy smokers suggesting that tobacco may be a confounding variable in the relation between alcohol and lung cancer. A recent meta-analysis among never smokers diagnosed with lung cancer suggested that alcohol does not play an independent role in lung cancer<sup>39</sup>.

High consumption of fruit and vegetables might protect against lung cancer. While consistent inverse associations between dietary intake of fruit and lung cancer risk have been reported in both case-control and cohort studies<sup>37</sup>, the role of vegetables is less clear. A review of several case-control studies have suggested a protective effect of high vegetable intake in lung cancer, but several prospective studies did not support this hypothesis<sup>40</sup>. Possible reasons of this discrepancy might result from bias including misclassification of exposure, retrospective dietary assessment and variability in food consumption between absence and presence of the disease. Also, residual confounding by cigarette smoking might explain these results, as some of the studies investigating the relation of vegetables with lung cancer did not adjust for smoking.

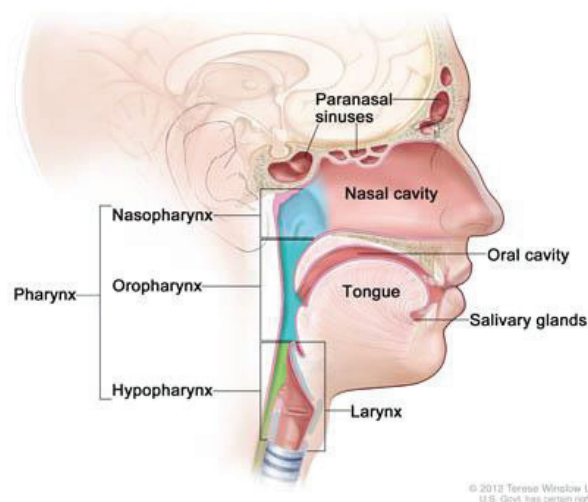
Several micronutrients have been investigated in relation to lung cancer and one of the most controversial is beta-carotene. Risk estimates from a pooled analysis of case-control and cohort studies reported lower lung cancer incidence for participants having high beta-carotene intake<sup>37</sup>. Thus, investigators hypothesised that beta-carotene may be related to lower lung cancer risk, and initiated randomized controlled trials (RCTs) were initiated in high lung cancer risk populations (middle-aged men that were smokers and/or exposed to asbestos). However, the Alpha-Tocopherol, Beta-Carotene, Prevention Cancer (ATBC) trial

reported that supplementation of beta-carotene increased lung cancer incidence<sup>37</sup>. An evaluation work from IARC concluded that high doses of beta-carotene supplements might not prevent lung cancer<sup>41</sup>. The World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) and several health authorities do not recommend that smokers take supplemental doses of beta-carotene. Several other micronutrient including vitamin A and E, lycopene, zinc and alpha-carotene have been investigated in relation to lung cancer risk, but to date evidence is not conclusive of any protective effect for lung cancer<sup>37 42</sup>.

## Head and neck cancer

### *Incidence and mortality*

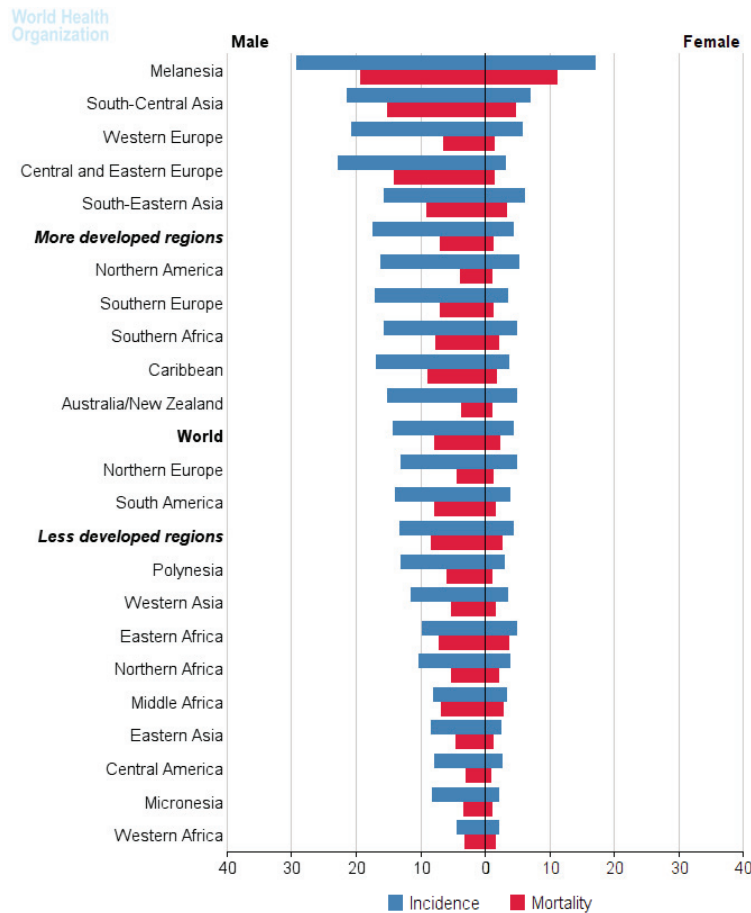
Head and neck cancer (HNC) encompass a group of tumours, at various sites around oral cavity, larynx, oropharynx and hypopharynx<sup>43 44</sup>. Over 90% of head and neck cancers are squamous cell carcinoma (HNSCC), arising from the epithelial cells that line the mucosal surfaces of the head and neck<sup>45</sup> (see [Figure 4](#)).



**Figure 4.** Schematic of various head and neck malignancies. Source from Terese Winslow LLC.

HNC is the sixth most common cancer type worldwide and affects nearly 650,000 individuals per year, and causes over 350,000 deaths<sup>46 47</sup>. The geographical inequalities in the burden of HNC are noticeable, with over 50 % of all cases occurring in developing countries<sup>15</sup>.

According to world estimations (see Figure 5), incidence rates of HNC are highest in Melanesia and South-Central Asia, in both sexes.



**Figure 5.** Age-standardized incidence and mortality rates (world) per 100,000 for head and neck cancer (lip, oral cavity, larynx, nasopharynx and other pharynx) in 2012. Source: Globocan 2012<sup>15</sup>.

Globally, men are affected significantly more by HNC than women with a ratio ranging from 2:1 to 4:1<sup>15</sup>. However, incidence by sex varies with anatomic location and has been changing as the number of female smokers has increased over the last decades. The male-female ratio is currently about 8:1 for laryngeal cancer and approximately 2:1 for nasopharyngeal cancer<sup>15</sup>.

Despite significant advances in multidisciplinary treatments of HNC including surgery, chemotherapy and radiation, the current 5-year survival rate from HNSCC of 50% has improved very little over the last 30 years<sup>48 49</sup>. Moreover, the prognosis is generally better for women compared to men and for neoplasms located in oral cavity and larynx than for those arising in the hypopharynx (see Figure 6).

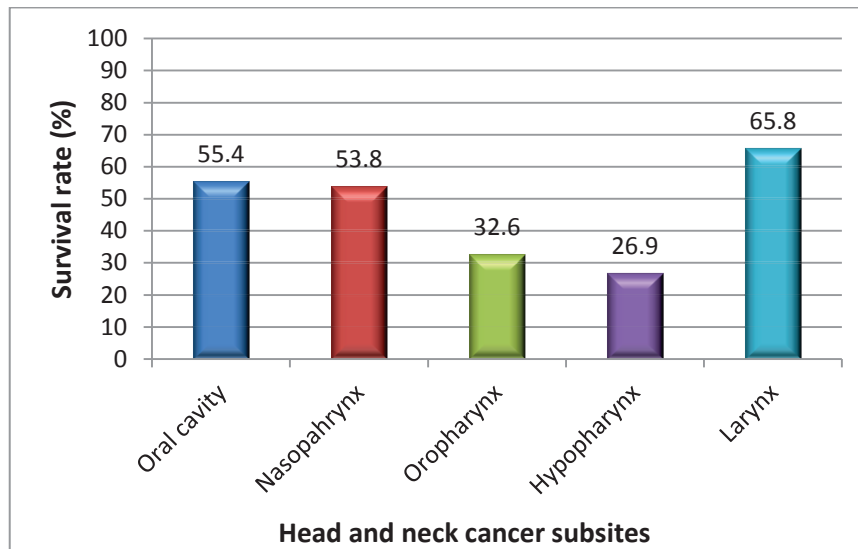


Figure 6. 5-year relative survival of HNC subsites in US, adapted from the SEER database<sup>49</sup>.

### **Risk factors of head and neck cancer**

Most HNC cases are attributable to tobacco smoking, alcohol drinking, and the established interaction between these two factors<sup>50</sup>. A positive association between tobacco smoking and HNC has been reported consistently in many studies<sup>1 27 50-52</sup>. Although the association with tobacco smoking is not as strong as with lung cancer, it remains very high with excess risk estimates among current smokers approaching 4 to 6-fold that among non-smokers<sup>1</sup>. Smoking is likely to be differentially associated with risk of subtypes of the HNC<sup>38 39 51 52</sup>. Both a recent meta-analysis of 28 case-control studies<sup>53</sup> and a large cohort study from Netherlands (NLCS)<sup>54</sup> reported that smoking was more strongly associated with laryngeal and oropharyngeal cancers than those from the oral cavity.

Similarly, substantial epidemiologic evidence from the last three decades demonstrates that alcohol consumption increases risk of the HNC. Recently, the Monograph Working Group of IARC concluded that there “*was sufficient evidence*” for the carcinogenicity of alcohol in animals and classified alcoholic beverages as carcinogenic to humans<sup>55</sup>. In particular, the group confirmed, and newly established, the causal link between alcohol consumption and several cancers including the cancer of the oral cavity, pharynx, larynx and oesophagus<sup>38</sup>.

Furthermore, several epidemiological studies showed a significant, positive, multiplicative interaction between tobacco smoking and alcohol consumption with HNC risk. As an example, the NLSC cohort reported that current smokers compared to never had higher risk



of HNC overall (relative risk [RR] 4.49, 95% CI 3.11-6.48) and that participants consuming  $\geq 30$  grams of alcohol per day compared with abstinence had also increased risk of HNC overall (RR = 2.74, 95% CI 1.85-4.06). When comparing participants smoking  $\geq 20$  cigarettes and drinking  $\geq 30$  g alcohol per day compared with never smokers abstaining from alcohol the RR was 8.28 (95% confidence interval [95% CI] 3.98-17.22 (interaction between categories of alcohol consumption and cigarette smoking was 0.03)<sup>54</sup>. Moreover, a positive association of tobacco smoking with HNC risk has been observed among non-drinkers and similarly alcohol consumption has been shown to increase risk of HNC among non-smokers<sup>56</sup>. Other studies reported similar results which aim to support the independent role of tobacco smoking and alcohol intake in HNC development.

Viral infection is a recognised risk factor for cancer of the head and neck. The association between Epstein-Barr virus infection and the development of head and neck cancer subtypes in particular nasopharyngeal cancer was first recognised in 1966<sup>57</sup>. More recently, human papilloma virus (HPV) has attracted attention<sup>58</sup>, with recent observational studies finding it to be a strong risk factor for the development of head and neck cancer. In particular, there is substantial evidence for a role of HPV16 in the oropharyngeal cancer etiology<sup>38 59</sup>.

Diet is estimated to contribute to about one-third of preventable cancers<sup>37 60</sup>. One of the most convincing pieces of epidemiological evidence for the role of dietary factors in cancer risk is the inverse relationship between the consumption of fruit and vegetables and HNC. Indeed, comprehensive evidence, both from case-control and cohort studies have shown that people with a low intake of fruit and vegetables have higher cancer incidence of the head and neck compared with those having a high fruit and vegetable intake<sup>37 61 62</sup>. Importantly, the last expert report from the World Cancer Research Fund (WCRF) stated that fruits, non-starchy vegetables, and foods containing carotenoids are probable decreased risk for HNC<sup>37</sup>. However, the evidence of the role of other food groups and the risk of HNC is inconsistent. Some epidemiological studies suggested that consumption of meat was associated with increased risk of HNC<sup>37 63 64</sup>, whereas other have reported no

association<sup>65 66</sup>. Despite the potential role of vegetable and fruit consumption in the etiology of HNC, residual confounding by smoking and alcohol is still a real concern.

While it can be difficult to disentangle the exact role of micronutrients individually with HNC, experimental and observational studies suggested that vitamin C<sup>37 67</sup>, vitamin E<sup>37 68</sup>, and beta-carotene<sup>37 69</sup> might reduce risk of HNC. Additionally, mainly retrospective studies have investigated folate and HNC risk, most reporting that higher dietary folate is inversely associated with HNC risk<sup>70</sup>. However, given the potential for reverse causality and recall bias to influence the results in retrospective case-control studies, it is difficult to interpret these results causally.

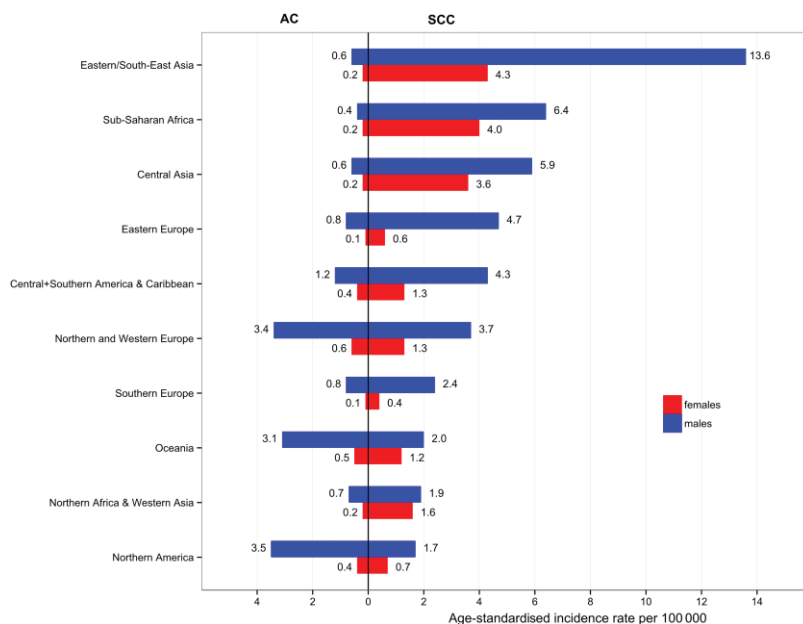
## **Esophageal cancer**

### ***Incidence and mortality***

Esophageal cancer is the eighth most common cancer worldwide, with an estimated 450,000 new cases in 2012, and the sixth most common cause of cancer death with an estimated 400,000 deaths<sup>15</sup>. Cancer of the esophagus includes mostly two histologic types, esophagus squamous cell carcinoma (ESCC) and esophagus adenocarcinoma (EAC), although ESCC is the predominant histologic type of esophageal cancer worldwide<sup>71</sup>. Throughout this dissertation we will focus on ESCC. In 2012, ESCC accounted for an estimated 398,000 new incident cases worldwide with a global ASR of 5.2 per 100,000 (7.7 in men and 2.8 in women, see [Figure 7](#)). The highest incidence rates are observed in Eastern/ South-eastern Asia (13.6 in men and 4.3 in women), sub-Saharan Africa (6.4 in men and 4.0 in women), and Central Asia (5.9 in men and 3.6 in women). The lowest incidence rates occurred in North America (1.7 in men and 0.7 in women), and Southern Europe (2.4 in men and 0.4 in women)<sup>71</sup>. Furthermore, it has been reported that about 80% of ESCC occurred within what is named the “esophageal cancer belt,” an area extends from northeast China to the Middle East<sup>72</sup>.

Globally, incidence rates of ESCC have been steadily declining in Western countries due to long-term reductions in alcohol intake and tobacco use in contrast to some Asian countries such as Taiwan where ESCC has been increasing, likely due to increase of tobacco and alcohol consumption<sup>73</sup>. It has been also reported that ESCC incidence vary across ethnicities

with two to three times higher estimates in blacks than in whites, in particular in North America<sup>74</sup>.



**Figure 7.** Age-standardised incidence rates per 100, 000 of oesophageal cancers by histological subtype, region and sex. AC, adenocarcinoma; SCC, squamous cell carcinoma. Source from Arnold M *et al*<sup>71</sup>.

There is generally no difference reported in survival between the two histological types, EAC and ESCC<sup>75</sup>. Geographical patterns of mortality are quite similar to those of incidence for both sexes due to the rapid course of the disease after diagnosis, then reflecting very poor survival, with 5-year survival rates of 19% (2003-2009) in the USA<sup>72</sup> and 12% (2000-2007) in Europe<sup>75</sup>.

**Risk factors of esophagus squamous cell carcinoma.**

The positive association between tobacco smoking and ESCC is well established. Cancer of the esophagus consistently occurs less frequently among participants who do not smoke, with supporting data from both control and cohort studies<sup>76-80</sup>. A recent prospective cohort study from Netherlands<sup>79</sup> reported that current smokers had higher risk of ESCC compared with never smokers, the RR being 2.63 (95% CI 1.47-4.69). However, the risk of ESCC among current smokers can be much stronger. Fredman *et al.* showed in a large prospective US

cohort of 476,606 participants that compared with non-smokers, current smokers were at increased risk for ESCC with a hazard ratio (HR) equal to 9.27 (95% CI 4.04-21.29)<sup>76</sup>.

Furthermore, the excess risk of ESCC among smokers relative to non-smokers depends on several aspects including smoking intensity and duration of smoking<sup>81</sup>. A number of studies also observed that smokeless tobacco products and betel liquid (with or without tobacco) increased risk of ESCC, in particular in Asia<sup>82-84</sup>.

The association between alcohol consumption and risk of ESCC has been described in many studies<sup>76-80 85</sup> and showed consistent positive association between alcohol consumption and ESCC risk. As an example, a recent meta-analysis of 40 case-control and 13 cohort studies showed that after adjusting for age, sex, and tobacco smoking the RR for the association between light alcohol drinking (12.5 g/d) and risk of ESCC was 1.38 (95% CI 1.14-1.67). The RRs were 2.62 (95% CI 2.07-3.31) among moderate drinkers (>12.5–<50 g/d) and 5.54 (95% CI 3.92-7.28) for high alcohol intake (50 g/d)<sup>86</sup>.

Tobacco smoking and alcohol intake act also synergistically (effect arising from the two factors) in a multiplicative manner on ESCC. As an example, a recent meta-analysis reported the effects of alcohol consumption and tobacco use on ESCC, both alone and in combination<sup>87</sup>. They found that never smokers with alcohol use had increased risk of ESCC (odds ratio [OR] being 1.21, 95% CI 0.81-1.81), although this association was statistically nonsignificant, and that never drinkers with tobacco use had also increase risk of ESCC (OR 1.36, 95% CI 1.14-1.61). This association was stronger when they looked both among alcohol and tobacco users with OR being 3.28 (95 % CI 2.11-5.08).

There have been numerous studies assessing the relation between diet and nutrition with ESCC etiology. Many case-controls and cohort studies showed that high consumption of fruit and vegetables were inversely associated with ESCC risk<sup>37 88-90</sup>. As an example, a recent prospective cohort study of 38,790 Japanese men aged 45–74 years showed that participants with highest consumption of total fruit and vegetables compared to the lowest had around 50% decreased risk of ESCC (HR 0.52, 95% CI 0.30-0.88)<sup>91</sup>.

A number of epidemiological studies have investigated the relation between micronutrients including folate, beta-carotene, vitamin C and vitamin E, and ESCC risk<sup>37 92-94</sup>, and some have proved an inverse association<sup>88 95</sup>. Folate intake is one of the micronutrient that has been assessed in several studies and showed consistent results. In 2006, Larson C *et al.* reported from a meta-analysis that the summary relative risk for the highest versus the lowest category of folate intake was 0.66 (95% CI 0.53-0.83) for ESCC<sup>94</sup>. Similarly, a recent large prospective study including 492 293 participants found that those with low levels of folate intake compared to high levels had around 90% increased risk of ESCC, the RR being 1.91 (95% CI 1.17-3.10)<sup>96</sup>.

Over the past two decades, the HPV in ESCC has received increasing attention and several studies using various methods have confirmed the presence of HPV in ESCC<sup>97-100</sup>. A recent case-control study from China showed that the presence of HPV in the esophagus was associated with increased risk of ESCC (OR 6.4, 95% CI 4.4-9.2)<sup>101</sup>.

Body mass index, a surrogate indicator of obesity, has been inversely associated with the risk of ESCC in some investigations<sup>37 102</sup>, although issues remain regarding confounding by smoking as alternative explanations. However, a recent population-based case-control study comprising 287 ESCC cases and 1,544 controls reported that after adjusting for smoking the risk of ESCC was reduced by 35% (range, 23%-44%) per 5-unit increase in recent BMI<sup>103</sup>.

## Kidney cancer

### ***Incidence and mortality***

As for cancer of the head and neck and lung, incidence rates of kidney cancer vary greatly across the world (see [Figure 8](#)). In 2012, the incidence rate of kidney cancer was the highest in Europe, with Czech Republic rising the top ASR worldwide per 100,000 for both sexes (24.1 in men and 10.5 in women)<sup>15</sup>. The lowest rates are observed in Asia (3.8 in men and 1.9 in women) and in Africa (1.4 in men and 1.1 in women).

In addition, stratifications by ethnic group as shown in United States indicate that African Americans and European Americans have on average two-fold higher ASR than Asian

Americans<sup>104</sup>. Figure 7 demonstrates also that men have approximately 2-fold higher kidney cancer incidence than women.

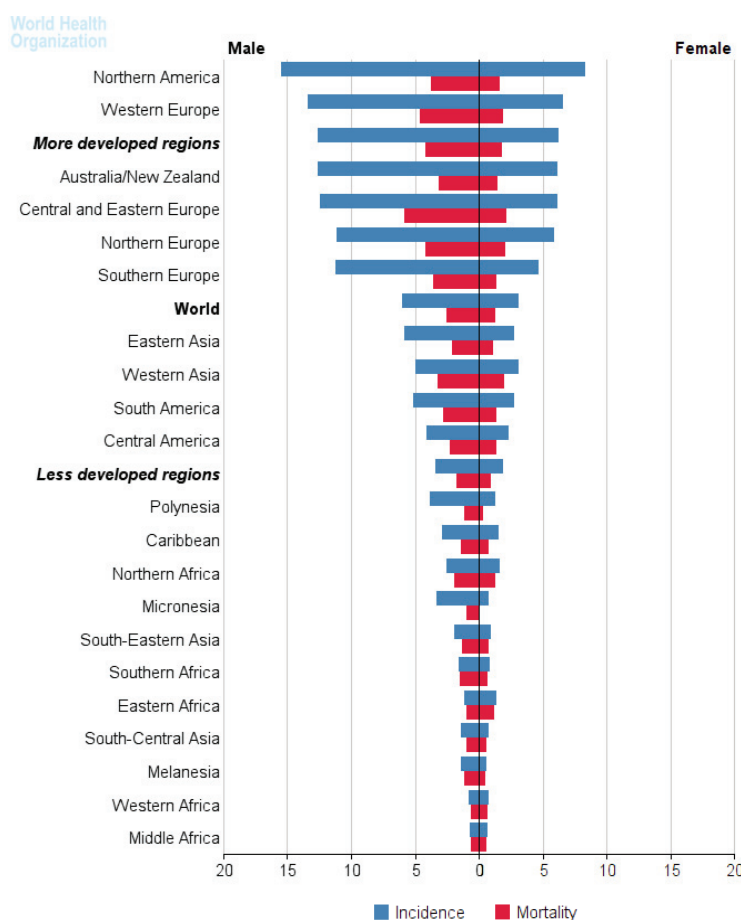
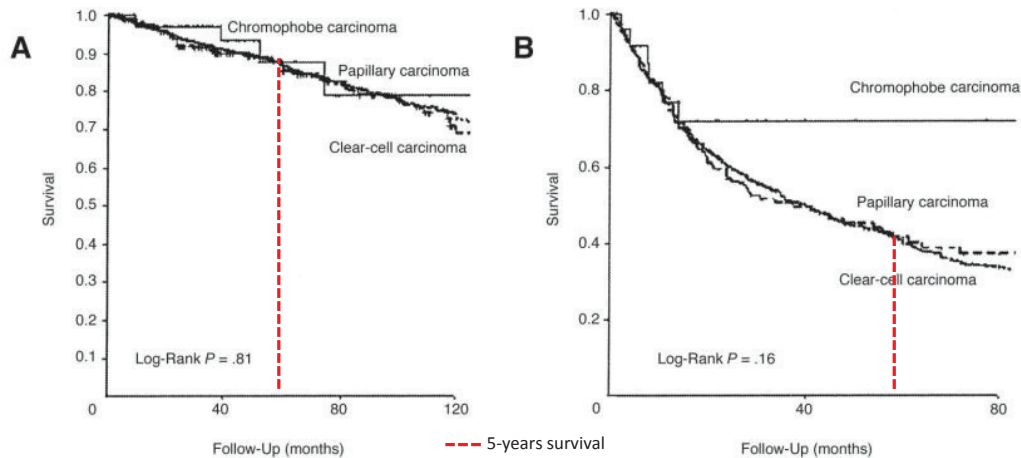


Figure 8. Age-standardized incidence and mortality rates (world) per 100,000 for kidney cancer in 2012. Source: Globocan 2012<sup>15</sup>.

Renal cell carcinoma (RCC) is the predominant type of kidney cancer and accounts for approximately 80 to 85% of cases<sup>104</sup>. Four subtypes of RCC have been established: papillary, chromophobe, collecting duct, and clear-cell, the latter accounting over 85% of RCC. Compared to the cancer of the lung and the head and neck, survival estimates after cancer diagnosis is better. From the Surveillance, Epidemiology and End Results Program (SEER)<sup>9</sup>, over 65% of kidney and renal pelvis cancer are diagnosed at the local stage (i.e. early stage including stage I, IIA and IIB), and 5-year survival for localized kidney and renal pelvis cancer is about 90% in United State. Further, recent data from eight international centres (see Figure 9) show that cases diagnosed with clear-cell carcinoma at early stages (i.e. stage I and

II) have a 5-year survival over 80% whereas cases diagnosed at a later stage (i.e. III and IV) have poorer prognosis with 5-year survival below 50%<sup>105</sup>.



**Figure 9.** Survival estimates in 4,063 patients with renal tumours according to histologic subtype and TNM stage. (A) TNM stage I to II. (B) TNM stage III to IV. Red dashed lines correspond to five-year survival. Source from Patard JJ et al.<sup>105</sup>.

### **Risk factors of kidney cancer**

Kidney cancer is less affected by smoking exposure than lung and head and neck. However, this exposure is considered as a causal risk factor for kidney cancer<sup>104 106</sup>. Tobacco smoking has been consistently reported to be a risk factor for this type of cancer, with increased risks compared to never smokers in the order of 50%<sup>17</sup>. Several epidemiological studies showed a clear dose-response pattern resulting in two to three-fold risk differences between heavy smokers and people who have never smoked<sup>17 107</sup>.

Obesity has been frequently studied in relation to cancer risk and different reports showed that this exposure increases the risk of kidney cancer<sup>37 108</sup>. In prospective studies conducted worldwide, overweight and obese individuals have an elevated risk of RCC in a dose-response manner<sup>109-111</sup>. Furthermore, hypertension has been consistently associated with higher risk of kidney cancer<sup>104 112</sup>. Several observational studies have reported an association with a history of long-term hypertension, and cohort studies of blood pressure measurements taken at baseline have generally shown a positive dose-response

relationship<sup>113-116</sup>. Despite the high correlation between obesity and hypertension, it has been suggested that these two risk factors alone may play an independent role in kidney cancer<sup>104 117</sup>. Furthermore, it has been pointed out that diabetes mellitus may increase kidney cancer risk, but results from case-controls as well cohort studies are not consistent<sup>104 118</sup>.

Despite alcohol intake being strongly positively related to cancer of the head and neck as previously shown, evidence from both case-control and cohort studies suggest that alcoholic drinks might protect against kidney cancer<sup>119 120</sup>. A recent meta-analysis of fifteen case-control studies observed an inverse association between alcohol consumption and RCC in both the overall alcohol intake group (OR 0.67, 95% CI 0.62-0.73) and subgroups analyses including sex, study design and geographical region<sup>123</sup>. A similar trend was observed in a meta-analysis of cohort studies involving over 750,000 subjects indicating that moderate alcohol consumption was associated with a lower risk (pooled multivariable RR 0.72, 95% CI 0.60-0.86) of RCC overall and this inverse association was also seen among both women and men<sup>124</sup>.

The evidence supporting dietary influences on kidney cancer etiology is less established than those observed for HNC. However, a recent pooled analysis of 13 prospective studies reported significant inverse associations between combined intake of fruit and vegetables in relation to renal cell carcinoma, with a stronger association for fruit<sup>125</sup>. Despite a potential role of fruit and vegetable intake in RCC etiology, it is difficult to disentangle which micronutrient might be involved in those associations. Different epidemiological studies investigating the role of beta-carotene, lycopene and folate in relation to kidney carcinogenesis are weak, but seems to point out an inverse association between both carotene<sup>125-127</sup> and vitamin E and kidney cancer<sup>128 129</sup>.



## **One-carbon metabolism: a candidate pathway**

As mentioned previously, dietary intake of fruit and vegetables has been associated with a potential decrease in risk among the three specific smoking-related cancer sites investigated in this dissertation. Fruit and vegetables are important dietary sources of B-vitamins and additional nutrients that are involved in the one-carbon pathway. Because these nutrients are important in maintaining DNA integrity and gene expression, they might have a potential role in preventing carcinogenesis, and present potential for cancer risk modification through changes in diet.

### **Overview of the one-carbon metabolism (OCM)**

OCM encompasses a series of cyclical biochemical reactions that involve the transfer of one carbon unit (see [Figure 10](#)). OCM pathways can be viewed several modular units that comprise in particular two interconnected cycles: the folate cycle and the methionine cycle.

#### *Folate cycle*

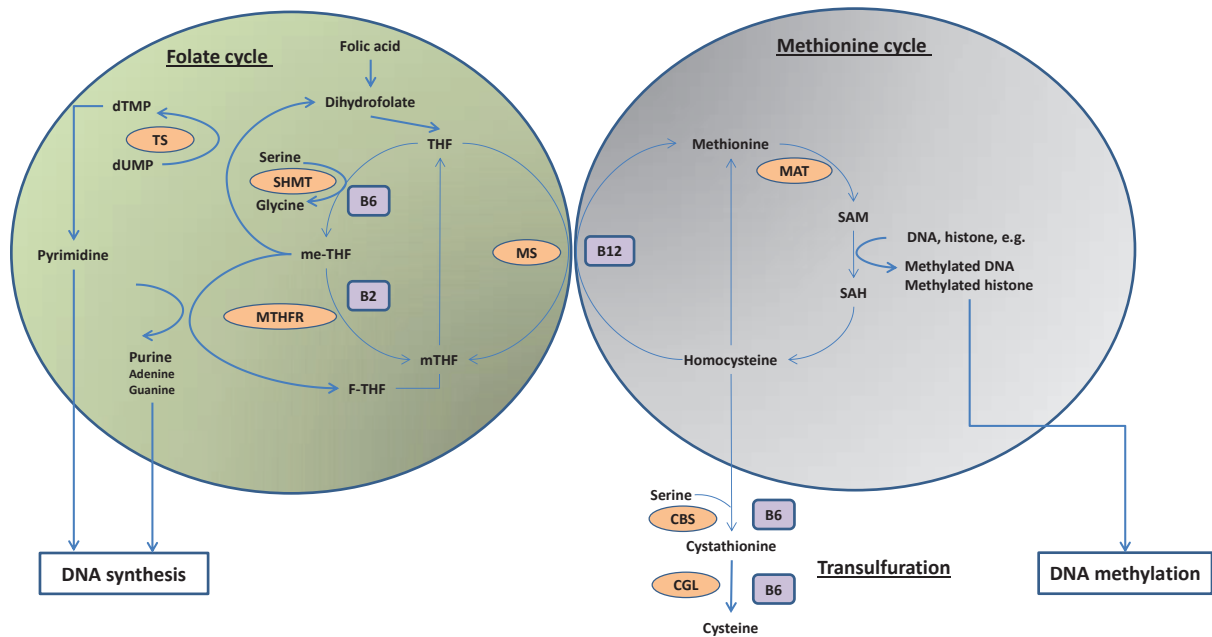
Folate plays a central role in OCM as a source of methyl groups. The term folate represents all naturally occurring forms of vitamin B9 in foods such as meat, fruits, vegetables, grain products, legumes, bread and seeds <sup>130</sup>, whereas folic acid is the synthetic form found in dietary supplements and fortified food.

In cells, folate either from dietary intake or folic acid, is reduced by a series of enzymatic reactions, leading to the generation of dihydrofolate and then tetrahydrofolate (THF) which is the most important carrier of OCM groups. THF is later converted to 5,10-methylene-THF (me-THF) by serine hydroxymethyl transferase (SHMT). me-THF is then either reduced to 5-methyltetrahydrofolate (mTHF) by methylenetetrahydrofolate reductase (MTHFR) or converted to 10-formyltetrahydrofolate (F-THF) through a sequence of steps. mTHF is demethylated to complete the folate cycle. After the demethylation of mTHF, the carbon is donated into the methionine cycle through the methylation of homocysteine to methionine by methionine synthase (MS) and its cofactor vitamin B12 (B12).

#### *Methionine cycle*

Methionine is an essential, amino acid and a precursor of S-adenosylmethionine (SAM), the primary methyl group donor for most biological methylations such as DNA, RNA, proteins, phospholipids, histones and neurotransmitters<sup>131 132</sup>. As shown in [Figure 10](#), the methionine

cycle begins with homocysteine that accepts the carbon from the folate pool through mTHF to generate methionine. Methionine, through methionine adenytransferase (MAT), is converted to S-adenosylmethionine (SAM), which is demethylated to form S-adenosylhomocysteine (SAH). SAH is finally converted back to homocysteine, completing the methionine cycle.



**Figure 10.** Overview of the OCM pathway. The OCM consists of the folate cycle (green) and the methionine cycle (gray). Cofactors are in purple and enzymes are in orange. SHMT: serine hydroxymethyltransferase, TS: thymidylate synthase, MTHFR: methylenetetrahydrofolate reductase, MAT: methionine adenytransferase, CBS: cystathionine-synthase, CGL: cystathionine  $\gamma$ -lyase, MS: methionine synthase, THF: tetrahydrofolate, me-THF: 5,10-methylenetetrahydrofolate, mTHF: 5-methyltetrahydrofolate, F-THF:10-formyltetrahydrofolate, dTMP: deoxythymidine monophosphate, dUMP: deoxyuridine monophosphate, SAH: S-adenosylhomocysteine, SAM: S-adenosylmethionine, B2: riboflavin, B6: pyridoxal 5'-phosphate, B12: cobalamin, adapted from Locasale et al., 2013<sup>133</sup>.

### Vitamin B6 influence

Vitamin B6 is abundant in food and its major sources come from whole grain, nuts, vegetables, bananas, milk and some types of meat<sup>130 134</sup>. In OCM, vitamin B6 is a co-factor of several enzymes (see Figure 10), including serine hydroxymethyltransferase (SHMT), cystathionine  $\beta$ -synthase (CBS) and cystathionine  $\gamma$ -lyase (CGL).

### **OCM and carcinogenesis**

Many epidemiological studies have investigated OCM nutrients in relation to cancer. While some studies reported significant associations, the evidence to date is not clear. The biological mechanisms behind the association between OCM and carcinogenesis is still evolving. In this dissertation, we introduce two primary mechanisms by which one-carbon metabolism has been proposed to influence cancer development: 1) damage in DNA repair and DNA synthesis and 2) influence on gene expression via aberrant methylation pattern.

### **DNA repair and DNA synthesis**

Folate is necessary for the *de novo* biosynthesis of purines and pyrimidines nucleotides, adequate of which is essential for genome stability<sup>135</sup>.

Purine nucleotides are necessary for DNA synthesis. In the process of the *de novo* biosynthesis of purines, the methyl group is provided by F-THF (see [Figure 10](#)) and then incorporated into the purine ring producing the following nucleotides adenine and guanine. Insufficient methyl groups from OCM factors imbalances may induce impaired DNA synthesis and repair. Synthesis of pyrimidines occurs when me-THF methylates deoxyuridylate monophosphate (dUMP) to form deoxythymidylate monophosphate (dTMP), a reaction essential for DNA replication and repair<sup>136</sup>.

It has been shown that mutations can occur either during the process of DNA repair, or during semiconservative DNA replication that leads to duplication of the cell genome<sup>137</sup>. In normal cells, a limited presence of folate reduces thymidylate from an imbalanced synthesis of dTMP (thymine) from dUMP (uracil) which leads in an increase in uracil content of DNA. This results in the incorporation of uracil into human DNA and which produces different types of genetic instability during DNA replication such as DNA strands breaks and chromosomal damage<sup>138</sup>. Thus, depletion of folate could not only block DNA repair capacity but also increase the need of DNA repair. In animal studies, this DNA and chromosomal damage has been shown to activate proto-oncogenes and inactivate tumour suppressor genes<sup>139</sup>, which are involved in cancer development. Knock *et al.* showed that low folate exposure increased intestinal tumorigenesis in mice through increased dUMP/dTMP ratio and DNA damage<sup>140</sup>. It has been also observed that a moderate dietary deficiency of folate increased incidence of colonic dysplasia and carcinoma in rats<sup>141</sup>.

While limited availability of folate in regular healthy cells could induce genetic instability, it has been suggested that folate may promote tumor progression after the neoplasm has been developed<sup>142</sup>. This controversial issue may be explained by the function of folate in nucleotide synthesis where tissues, including tumours, proliferate rapidly, and thus have an increased requirement for nucleotides. Then many cancers up-regulate folate receptors, and anti-folate drugs, such as 5-fluorouracil or methotrexate, are efficacious in treatment of cancer by blocking the DNA replication of cancer cells<sup>142 143</sup>.

In addition, vitamin B6 imbalance may disturb different reactions in the OCM resulting also in dysregulation of DNA synthesis and DNA methylation, which leads to cancer development. It has been reported that vitamin B6 deficient subjects are thought to be more susceptible to DNA and chromosome breaks by a similar mechanism as described above. In vitamin B6 deficiency, availability of me-THF is decreased, reducing the methylation of dUMP and thereby increasing misincorporation of uracil, causing genomic instability. *In vitro* studies showed growth inhibition by vitamin B6 on different type of cancers<sup>144 145</sup>, including hepatoma cells of rat<sup>146</sup>. Most human studies on the association of vitamin B6 with carcinogenesis focused on colorectal cancer and have suggested that deficiency of vitamin B6 may increase risk<sup>147</sup>. However, since vitamin B6 is not solely dependent to the OCM pathway, this nutrient may also suppress carcinogenesis by reducing cell proliferation, angiogenesis, oxidative stress, inflammation, and nitric oxide synthesis<sup>148-150</sup>.

### **DNA methylation**

DNA methylation is an important epigenetic modification in mammals that confers an adequate regulation of gene expression, embryonic development and chromosome stability<sup>135 151 152</sup>. Epigenetics is defined as “the study of heritable changes in gene expression that occur without a change in DNA sequence”<sup>153</sup>, and unlike genetics, epigenetic changes are dynamic and reversible. In humans, DNA methylation is a covalent chemical modification, resulting in the addition of a methyl (-CH<sub>3</sub>) group at the carbon 5 position of the cytosine ring. This chemical modification can occur at millions of locations on the DNA sequence and occurs most of the time at CpG sites<sup>151</sup>. Between 70 to 80% of CpG

dinucleotides are methylated in mammals. CpG sites are generally dispersed in the genome non-uniformly, with the exception of short regions that are grouped in clusters and called 'CpG islands'<sup>152</sup>. These CpG islands rich in CpG are observed mostly in the promoter region of many genes and are in general unmethylated, whereas CpG sites dispersed non-uniformly are usually methylated under regular conditions. When methylation of promoter CpG islands occurs, transcription is inhibited and gene expression silenced<sup>152</sup>.

### **DNA methylation and cancer**

Adequate DNA methylation is crucial for cell development and function, thus any anomalies in this process may lead to diverse diseases, including cancer. In fact, cells in cancer display a different methylome compared to normal cells. Interestingly, a decrease in genomic methylation at dispersed CpG sites and abnormal hypermethylation in gene promoter CpG islands are often observed in human cancers. The hypermethylation phenomena may contribute to silencing of tumor suppressors, whilst hypomethylation may activate oncogenes<sup>154</sup>.

Regional increases in DNA methylation concentrations have been reported from many studies. Many genes have been found to be hypermethylated in cancer including *p16*, *BRCA1* and *RASSF1A*<sup>155</sup>. Hypermethylation was reported in many types of human cancers including breast<sup>156 157</sup>, prostate<sup>158</sup>, leukemia<sup>159 160</sup> and lung<sup>161</sup>.

Global loss of DNA methylation of the genome, termed hypomethylation, has been suggested to initiate carcinogenesis. Hypomethylation is observed among several malignancies<sup>162 163</sup> including solid tumours such as prostate tumours<sup>164</sup>, metastatic hepatocellular cancers<sup>165</sup>, cervical cancers<sup>163</sup>, and also in haematological malignancies including leukaemia<sup>166</sup>.

DNA methylation levels may be tissue-specific because the variation among different tissues is substantial<sup>131 167</sup>. Some studies reported that DNA methylation levels in carcinogenic tissues did not differ from tissues characterised as benign or normal tissues surrounding carcinogenic tissues<sup>168 169</sup>.

### **OCM and methylation**

Many nutrients involved in the OCM including vitamin B6 (SHMT activity), vitamin B2 (MTHFR stability) and vitamin B12 (MS function)<sup>170</sup> are important to maintain the one carbon flux, and deficiencies or imbalances of these biomarkers may lead to aberrant DNA methylation and subsequently carcinogenesis.

The OCM, and thus DNA methylation, operate under the control of multiple influences. Moderate elevations in plasma homocysteine concentration have been reported to be associated with increased concentration of SAH, but not SAM, and increased SAH concentration has been associated with global DNA hypomethylation<sup>171</sup>. For example, several *in vitro* studies demonstrated a clear association between folate deficiency and DNA stability where it has been shown that deficiency of folate decreases genomic DNA methylation<sup>132</sup>. However, studies in humans have been inconclusive. Some suggest that altered folate status can affect global DNA methylation in humans<sup>172 173</sup> whereas others reports no significant changes in DNA methylation in relation to folate concentrations<sup>174 175</sup>. In addition, it has been suggested that in vitamin B6 deficiency, DNA methylation is decreased because of less availability of SAM, the methyl group donor. Some studies have shown an association between B6 and hypermethylation<sup>176 177</sup>, but to date no studies have investigated the relation of hypomethylation with B6 deficiency.

One-carbon pathway is a complex mechanism involving many factors. With an aim to further understand its relation in cancer, we conducted studies with a specific focus on some of the OCM cornerstones such as the B-vitamins involved in the folate cycle as well as the methionine and homocysteine engaged in methionine cycle. For clarity, OCM nutrients, factors or biomarkers refers in this dissertation to the combination of vitamin B2, vitamin B6, folate, vitamin B12, homocysteine and methionine.

## Nutrients of the candidate pathway and smoking-related cancer

### Overview

Among OCM biomarkers, folate is by far the most studied analyte in relation to cancer. Over the past decades, most human studies assessing the role of folate (measurement from dietary intake or blood) with carcinogenesis focused on colorectal cancer (CRC). The accumulative of epidemiologic and clinical evidence have reported about 20% to 40% decreased risk of CRC or adenoma in subjects with the highest dietary intake of folate compared to those being in the lowest intake, although the evidence is less consistent for circulating levels of folate<sup>178-180</sup>.

### OCM analytes and lung cancer

Among OCM biomarkers, folate is also by far the most studied analyte in relation to lung cancer as previously shown for CRC. However, data on the effect of folate on the subsequent incidence of lung cancer are limited and inconclusive. Table 1 summarizes the literature on the selected OCM nutrients in relation to lung cancer risk. Inverse association of dietary folate intake with lung cancer risk have been reported in few case-control studies<sup>181 182</sup>. In contrast, a recent meta-analysis of nine cohort studies found a very small or no association with lung cancer risk<sup>183</sup>, and only one of the included studies alone reported a significant inverse association<sup>184</sup>. Dietary intake of other micronutrients related to the OCM pathway has been studied in relation to lung cancer development, including vitamin B2, B6, B12 and methionine, but no significant associations were found in those studies (see Table 1).

Investigations on the relation between circulating concentrations of OCM factors with lung cancer risk have been assessed very sparsely in a prospective setting. The ATBC study was the first to evaluate whether pre-diagnostic circulating of B-vitamins (B6, B9 and B12) and homocysteine were associated with lung cancer risk. This nested case-control study of 300 lung cancer cases and matched controls among Finnish male smokers did not find any clear association between circulating concentrations of folate, vitamin B12 and homocysteine and lung cancer risk<sup>185</sup>. However, subjects in the top quintiles of circulating concentrations of vitamin B6 had a two-fold decrease risk of lung cancer compared with those in the bottom

quintiles (OR 0.51; 95% CI 0.23-0.93). Ten years after the ATBC study on lung cancer, a nested case-control study of 899 cases and 1,770 controls from the EPIC cohort reported strong inverse associations between circulating concentrations of vitamin B6 and methionine independently to smoking status with lung cancer risk, resulting in 2.5-fold risk differences between the top and bottom 25% of the population<sup>13</sup>. Moreover, this study reported significant inverse association between circulating concentrations of folate and lung cancer risk. Overall, it remains unclear if factors of OCM pathway are important in lung cancer development.

**Table 1.** Associations between OCM factors (vitamin B2, B6, B12, folate, methionine and homocysteine) and lung cancer risk

Author	Year	Exposure	Studies	Subjects	Results <sup>a</sup>
<b>Dietary assessment</b>					
Yu-Fei Zhang <i>et al.</i> <sup>183</sup>	2014	dietary folate intake	meta-analysis	9 cohort studies	High vs low dietary folate intake: RR = 0.92(0.84-1.01), <i>p</i> for trend= 0.08 Per 100ug/day increment in folate lung cancer RR= 0.99(0.97-1.01), <i>p</i> for trend 0.31
Yumie Takata <i>et al.</i> <sup>186</sup>	2013	dietary folate intake	prospective cohort study	359 lung cancer cases (men)	highest vs. lowest quartile of dietary folate intake: HR=0.99 (0.70 – 1.40), <i>p</i> for trend= 0.92
Dai WM <i>et al.</i> <sup>181</sup>	2013	dietary folate intake	meta-analysis	4,390 lung cancer cases and 6,138 controls	Overall inverse association of dietary foalte intake: OR=0.74 (0.65-0.84), <i>p</i> for trend <0.001
Yumie Takata <i>et al.</i> <sup>187</sup>	2012	dietary intake of B2, B6, folate, B12 and methionine	prospective cohort study	359 lung cancer cases (women)	highest vs. lowest quartile of dietary B2 intake: HR=0.62 (0.43-0.89), <i>p</i> for trend =0.03 highest vs. lowest quartile of dietary B6 intake: HR=1.21 (0.84-1.74), <i>p</i> for trend =0.44 highest vs. lowest quartile of dietary folate intake: HR=0.96 (0.70 – 1.32), <i>p</i> for trend = 0.79 highest vs. lowest quartile of dietary B12 intake: HR=0.76 (0.55-1.05), <i>p</i> for trend =0.19 highest vs. lowest quartile of dietary methionine intake: HR=0.87 (0.58-1.29), <i>p</i> for trend =0.26



Basset JK <i>et al.</i> <sup>188</sup>	2012	dietary intake of B2, B6, folate, B12 and methionine	prospective cohort study	348 lung cancer cases	highest vs. lowest quintile of dietary B2 intake: HR=0.79 (0.55-1.13), <i>p</i> for trend=0.15 highest vs. lowest quintile of dietary B6 intake: HR=1.03 (0.74-1.43), <i>p</i> for trend=0.67 highest vs. lowest quintile of dietary folate intake: HR=1.02 (0.69 – 1.50), <i>p</i> for trend = 0.60 highest vs. lowest quintile of dietary B12 intake: HR=1.12 (0.79-1.59), <i>p</i> for trend=0.52 highest vs. lowest quintile of dietary methionine: intake HR=0.93 (0.66-1.31), <i>p</i> for trend=0.90
Nina Roswall <i>et al.</i> <sup>189</sup>	2010	dietary folate intake	prospective cohort study	721 lung cancer cases	highest vs. lowest quartile of dietary folate intake: RR=0.99 (0.79-1.23), <i>p</i> for trend=0.68
Christopher G Slatore <i>et al.</i> <sup>190</sup>	2008	dietary folate intake	prospective cohort study	521 lung cancer cases	highest vs. lowest tertile of dietary folate intake: HR=0.99 (0.79-1.23), <i>p</i> for trend=0.68
GC Kabat <i>et al.</i> <sup>191</sup>	2008	dietary intake of B2, B6, folate, B12 and methionine	prospective cohort study	358 lung cancer cases (women)	highest vs. lowest quintile of dietary B2 intake: RR=1.30 (0.94-1.80), <i>p</i> for trend=0.08 highest vs. lowest quintile of dietary folate intake: RR=1.12 (0.83 – 1.52), <i>p</i> for trend = 0.43 highest vs. lowest quintile of dietary methionine: intake RR=0.95 (0.69-1.31), <i>p</i> for trend=0.76
Eunyoung Cho <i>et al.</i> <sup>192</sup>	2006	dietary folate intake Totale folate intake	pooled analysis (8 prospectives studies)	3,206 lung cancer cases	highest vs. lowest quintiles of dietary folate intake: RR=0.88 (0.74-1.04), <i>p</i> for trend =0.08 highest vs. lowest quintiles of total folate intake: RR=1.02 (0.83-1.26), <i>p</i> for trend =0.51
Hongbing Shen <i>et al.</i> <sup>182</sup>	2003	dietary folate intake	case-control study	470 lung cancer cases and 472 matched controls (former smokers)	dietary folate intake: OR=0.53 (0.35 –0.80) (highest vs. lowest quartile), <i>p</i> for trend <0.001
Jian-Min Yuan <i>et al.</i> <sup>193</sup>	2003	dietary folate intake	prospective cohort study	482 lung cancer cases	highest vs. lowest quintile of dietary folate intake: RR=0.82 (0.60 – 1.11)
Laura E. Voorrips <i>et al.</i> <sup>194</sup>	2000	dietary folate intake	prospective cohort study	939 lung cancer cases (men)	highest vs. lowest quintile of dietary folate intake: RR=0.83 (0.39 – 1.75), <i>p</i> for trend = 0.03
Elisa V. Bandera <i>et al.</i> <sup>184</sup>	1997	dietary folate intake	prospective cohort study	395 lung cancer cases	highest vs. lowest tertile of dietary folate intake: RR=0.70 (0.55 – 0.89), <i>p</i> for trend<0.001
<b>Circulating concentrations</b>					

Terry J. Hartamn <i>et al.</i> <sup>185</sup>	2001	Serum of vitamin B6, folate and B12	nested case-control	300 lung cancer cases and matched controls (male smokers)	highest vs. lowest quintile of serum B6: OR=0.51 (0.23-0.93), <i>p</i> for trend =0.02 highest vs. lowest quintile of serum folate: OR=0.96 (0.52-1.72), <i>p</i> for trend =0.28 highest vs. lowest quintile of serum B12: OR=1.41 (0.80-2.50), <i>p</i> for trend =0.14 highest vs. lowest quintile of serum homocysteine: OR=0.61 (0.32-1.17), <i>p</i> for trend =0.41
Mattias Johansson <i>et al.</i> <sup>13</sup>	2010	Serum of vitamin B2, B6, folate, B12, methionine and homocysteine	nested case-control	899 lung cancer cases and 1,770 matched controls	highest vs. lowest quartile of serum B6: OR=0.44 (0.33-0.60), <i>p</i> for trend <0.001 highest vs. lowest quartile of serum methionine: OR=0.52 (0.39-0.69), <i>p</i> for trend <0.001 highest vs. lowest quartile of serum folate: OR=0.68 (0.51-0.90), <i>p</i> for trend =0.001 highest vs. lowest quartile of serum B12: OR=1.35 (1.00-1.82), <i>p</i> for trend = 0.04 highest vs. lowest quartile of serum B2: OR=0.99 (0.73-1.35), <i>p</i> for trend =0.81 highest vs. lowest quartile of serum homocysteine: OR=0.76 (0.54-1.07), <i>p</i> for trend =0.62
<b>Vitamin supplementation</b>					
Stein Emil Vollset <i>et al.</i> <sup>195</sup>	2013	folic acid supplementation	meta-analysis (13 RCTs*)	272 lung cancers	Overall folic acid vs. Non-folic acid group: RR= 1.08 (0.86 – 1.35), <i>p</i> for trend = 0.37
Marta Ebbing <i>et al.</i> <sup>196</sup>	2009	folic acid and B6 supplementation	meta-analysis (2 RCTs*)	92 lung cancers	Overall folic acid vs. Non-folic acid group: HR= 1.59 (0.92 – 2.75) Overall vitamin B6 vs. Non-vitamin B6 group: HR= 1.06 (0.62 – 1.82)

\*RCTs= randomized clinical trials

<sup>a</sup> Multivariate-adjusted OR,RR or HR and 95% confidence intervals (CIs) in parentheses

### OCM factors and head and neck cancer

While little evidence have emerged on food groups linked to specific cancer sites, cancer of the head and neck consistently occurs less frequently among subject with high consumption of vegetables and fruit, with supporting data from both case-control<sup>62</sup> and prospective studies<sup>61 197</sup>. To date, very few epidemiological studies - mainly case-control studies - assessed the association between B-vitamins and HNC. Results of these studies are summarized in Table 2. Overall, some case-control studies reported no or little relationship between intake of folate and vitamin B6 with HNC risk<sup>92 198 199</sup> where as other studies

indicated a significant inverse association<sup>69 200-202</sup>. Furthermore, a recent pooled analysis of individual data from ten case-control studies participating in the International Head and Neck Cancer Epidemiology (INHANCE) Consortium and including over 5,000 cases and 13,000 controls, reported that high levels of folate intake compared to those with low levels were found to be inversely associated with overall oropharyngeal cancer risk<sup>70</sup>. The other OCM nutrients measured by dietary intake, including vitamin B2, B12 and methionine, did not reveal any clear association with HNC risk.

Only a few retrospective case-control studies have reported on circulating OCM biomarkers and HNC risk, most indicating lower folate concentrations and higher homocysteine concentrations among cases than among controls<sup>203-206</sup>. However, recall bias and reverse causality is particularly a concern because the nutrient (from dietary questionnaire or blood measurement) is assessed after diagnosis for cancer cases and thus, it is difficult to interpret those studies in the context of cancer etiology.

**Table 2.** Associations between OCM factors (vitamin B2, B6, B12, folate, methionine and homocysteine) and the head and neck cancer risk

Author	Year	Exposure	Studies	Subjects	Results <sup>a</sup>
<b>Dietary assessment</b>					
Carlotta Galeone <i>et al.</i> <sup>70</sup>	2014	dietary intake of folate	case-control	5,127 cancers of oral cavity and pharynx and 13,249 controls	highest vs. lowest quintile of dietary folate intake: OR=0.65 (0.43- 0.99), <i>p</i> for trend =0.04
Keitaro Matsuo <i>et al.</i> <sup>201</sup>	2012	dietary intake of folate	case-control	409 HNC cases	highest vs. lowest quartile of dietary folate intake: OR=0.53 (0.36- 0.77), <i>p</i> for trend <0.001
D. Aune <i>et al.</i> <sup>92</sup>	2011	dietary folate intake	case-control	798 UADT* cancer cases and 2,032 controls 283 cancer cases of oral cavity and pharynx 281 larynx cancers	highest vs. lowest quartile of dietary folate intake: OR=0.41 (0.26 – 0.65), <i>p</i> for trend <0.0001 highest vs. lowest quartile of dietary folate intake: OR=0.49 (0.24 – 0.98), <i>p</i> for trend =0.04 highest vs. lowest quartile of dietary folate intake: OR=0.54 (0.27 – 1.10), <i>p</i> for trend =0.07
C. Pelucchi <i>et al.</i> <sup>202</sup>	2003	dietary intake of folate and methionine	case-control	749 cancer cases of the oral cavity and pharynx and 1,772 controls	highest vs. lowest tertile of dietary folate intake: OR=0.53 (0.40-0.69), <i>p</i> for trend <0.001 highest vs. lowest tertile of dietary methionine intake: OR=1.14 (0.87- 1.50), <i>p</i> for trend =0.33

Ettore Bidoli <i>et al.</i> <sup>200</sup>	2003	dietary intake of B2, B6 and folate	case-control	527 laryngeal cancer cases and 1,297 controls	highest vs. lowest quintile of dietary B2 intake: OR=1.00 (0.70- 1.50), <i>p</i> for trend =0.05 highest vs. lowest quintile of dietary B6 intake: OR=0.60 (0.40- 0.90), <i>p</i> for trend <0.05 highest vs. lowest quintile of dietary folate intake: OR=0.40 (0.20- 0.60), <i>p</i> for trend <0.01
Stephanie J. Weinstein <i>et al.</i> <sup>199</sup>	2002	dietary intake of folate and methionine	case-control	341 oral cancers cases and 521 controls	highest vs. lowest quartile of dietary folate intake: OR=0.60 (0.40-1.00), <i>p</i> for trend =0.05 highest vs. lowest quartile of dietary methionine intake: OR=1.40 (0.90-2.20), <i>p</i> for trend =0.21
Eva Negri <i>et al.</i> <sup>69</sup>	2000	dietary intake of B2, B6 and folate	case-control	754 oral cancer cases and 1,775 controls	highest vs. lowest tertile of dietary B2 intake OR=0.64 (0.40-1.01), <i>p</i> for trend =0.09 highest vs. lowest tertile of dietary B6 intake OR=0.27 (0.15-0.50), <i>p</i> for trend <0.0001 highest vs. lowest tertile of dietary folate intake: OR=0.32 (0.20 – 0.52), <i>p</i> for trend <0.0001
E. De Stefani <i>et al.</i> <sup>198</sup>	1999	dietary intake of B2, B6, folate and B12	case-control	133 UADT* cancer cases and 383 controls	highest vs. lowest tertile of dietary B2 intake: OR=2.00 (1.10-3.5), <i>p</i> for trend =0.04 highest vs. lowest tertile of dietary B6 intake: OR=0.80 (0.50-1.40), <i>p</i> for trend =0.16 highest vs. lowest tertile of dietary folate intake: OR=0.90 (0.50 – 1.60), <i>p</i> for trend = 0.90 highest vs. lowest tertile of dietary B12 intake: OR=1.80 (1.00-3.2), <i>p</i> for trend =0.05

#### Vitamin supplementation

Stein Emil Vollset <i>et al.</i> <sup>195</sup>	2013	folic acid supplementation	meta-analysis (13 RCTs**)	29 cancers of the lip, mouth and pharynx	Overall folic acid vs. Non-folic acid group RR= 1.38 (0.66 – 2.86), <i>p</i> for trend = 0.26
---	------	----------------------------	---------------------------	--	---

\*UADT =upper aero digestive tract; RCTs\*\*= randomized clinical trials

<sup>a</sup> Multivariate-adjusted OR or RR and 95% confidence intervals (CIs) in parentheses

#### OCM factors and esophageal squamous cell carcinoma

Evidence on the association between the selected OCM micronutrients in this dissertation and cancer risk of the squamous cell esophagus has been limited and is summarized in Table 3. Globally, most of epidemiological studies were case-control studies and assessed the association between folate intake and ESCC risk. Three independent retrospective case-control studies<sup>93 207 208</sup> did not find any clear association between folate intake and risk of

ESCC. Surprisingly, a meta-analysis comprising four case-control studies showed a significant inverse association between folate intake and ESCC risk with a summary relative risk when comparing the highest versus the lowest category of folate intake being 0.66 (95% CI 0.53-0.83)<sup>94</sup>. Similarly, a recent large prospective study including 492 293 participants found that those with low levels of folate intake compared to high levels had around 90% increased risk of ESCC, the relative risk being 1.91 (95% CI 1.17-3.10)<sup>96</sup>. We did not find any study assessing the relation of circulating concentrations of B-vitamins and related metabolites of the OCM with ESCC.

**Table 3.** Associations between OCM biomarkers factors (vitamin B2, B6, folate and methionine) and esophageal squamous cell carcinoma risk

Author	Year	Exposure	Studies	Subjects	Results <sup>a</sup>
Linda M. Brown <i>et al.</i> <sup>207</sup>	1998	dietary intake of B2 and folate	case-control	114 ESCC cases and 681 controls (whites)  219 ESCC cases and 557 controls (blacks)	highest vs. lowest quartile of dietary B2 intake OR=0.90, <i>p</i> for trend =0.42 highest vs. lowest quartile of dietary folate intake OR=0.70, <i>p</i> for trend =0.63  highest vs. lowest quartile of dietary B2 intake OR=1.20, <i>p</i> for trend =0.67 highest vs. lowest quartile of dietary folate intake OR=1.10, <i>p</i> for trend =0.85
Susan T. Mayne <i>et al.</i> <sup>208</sup>	2001	dietary folate, intake	case-control	206 ESCC cases and 687 controls	75th vs. 25th of dietary B2 intake OR=1.26 (0.84-1.89) 75th vs. 25th of dietary B6 intake OR=0.45 (0.30-0.69) 75th vs. 25th of dietary folate intake OR =0.58 (0.39-0.86) 75th vs. 25th of dietary B12 intake OR=1.51 (1.15-2.00)
C. Galeone <i>et al.</i> <sup>93</sup>	2005	dietary intake of B6, folate and methionine	case-control	351 ESCC cases and 875 controls	highest vs. lowest tertile of dietary B6 intake OR=0.88 (0.60-1.31), <i>p</i> for trend =0.55 highest vs. lowest tertile of dietary folate intake OR=0.68 (0.46-1.00), <i>p</i> for trend =0.05 highest vs. lowest tertile of dietary methionine intake OR=1.11 (0.76-1.62), <i>p</i> for trend =0.58
Susanna C. Larsson <i>et al.</i> <sup>94</sup>	2006	dietary intake of folate	meta-analysis (4 case-control studies)	929 ESCC cases	highest vs. lowest category of dietary folate intake OR=0.66 (0.53- 0.83)

Q Xiao <i>et al.</i> <sup>96</sup>	2014	dietary intake of B6, folic acid, B12 and methionine	prospective cohort	185 ESCC cases	1th quintiles vs. 3rd quintile of dietary B6 intake: RR=1.38 (0.91 – 2.12), <i>p</i> for trend =0.01 1th quintiles vs. 3rd quintile of dietary folate intake: RR=1.91 (1.17 – 3.10), <i>p</i> for trend =0.02 1th quintiles vs. 3rd quintile of dietary B12 intake: RR=1.21 (0.78 – 1.87), <i>p</i> for trend =0.13 1th quintiles vs. 3rd quintile of dietary methionine intake: RR=1.45 (0.91 – 2.31), <i>p</i> for trend =0.16
---------------------------------------	------	---	-----------------------	----------------	---

ESCC= esophagus squamous cell carcinoma

<sup>a</sup> Multivariate-adjusted OR or RR and 95% confidence intervals (CIs) in parentheses

### OCM factors and kidney cancer

Although the relation between fruit and vegetable intake with kidney cancer is intriguing and consistently observed in both retrospective and prospective case-control studies, the associations between micronutrients from dietary intake related to the OCM with kidney cancer are few and inconsistent. To date, only four studies (two retrospective case-control studies and two prospective cohorts) have investigated intake of OCM nutrients, including vitamin B2, B6, B12, folate and methionine, in relation to risk of kidney cancer (see Table 3). As a whole, these studies found little evidence that higher intakes of nutrients related to OCM lower risks of RCC and concluded that those OCM nutrients may have little influence on kidney carcinogenesis<sup>209-211</sup>.

Circulating concentrations of B-vitamins have been assessed in relation to multiple cancers, but only one study with a prospective design has been published regarding RCC to date<sup>212</sup>. The ATBC study investigated the association between serum OCM biomarkers (folate, vitamin B2, B6, B12, cysteine and homocysteine) and risk of RCC. This nested case-control study of 224 case-control pairs did not find any clear association between those biomarkers and RCC risk. A role of OCM in renal cell development is biologically plausible, but evidence regarding such a role is limited.

**Table 4.** Associations between OCM factors (vitamin B2, B6, B12, folate, methionine and homocysteine) and kidney cancer risk

Author	Year	Exposure	Studies	Subjects	Results <sup>a</sup>
<b>Dietary assessment</b>					
Kaye E. Brock <i>et al.</i> <sup>210</sup>	2012	dietary folate intake	case-control	323 RCC* cases and 1,827 matched controls	highest vs. lowest quartile of dietary folate intake: OR=0.80 (0.50 – 1.20), <i>p</i> for trend= 0.12
Christina Bosseti <i>et al.</i> <sup>209</sup>	2006	dietary intake of B2, B6 and folate	case-control	767 RCC* cases and 1,534 matched controls	highest vs. lowest quintile of dietary B2 intake: OR=0.88 (0.66-1.19) highest vs. lowest quintile of dietary B6 intake: OR=0.85 (0.64-1.13) highest vs. lowest quintile of dietary folate intake: OR=0.85 (0.64 – 1.12)
Eunyoung Cho <i>et al.</i> <sup>211</sup>	2013	Total intake of B6, folate, B12 and methionine	pooled analysis (2 prospective studies)	436 RCC* cases	highest vs. lowest quintile of total B6 intake: RR=0.88 (0.65-1.19), <i>p</i> for trend=0.55 highest vs. lowest quintile of total folate intake: RR=0.84 (0.60 – 1.18), <i>p</i> for trend = 0.52 highest vs. lowest quintile of total B12 intake: RR=1.24 (0.90-1.70), <i>p</i> for trend=0.61 highest vs. lowest quintile of total methionine: intake RR=1.29 (0.93-1.78), <i>p</i> for trend=0.10
<b>Circulating concentrations</b>					
Todd M. Gibson <i>et al.</i> <sup>212</sup>	2010	Serum of vitamin,B2, B6, folate, B12, and homocysteine	nested case-control	224 RCC* cases and 224 matched controls	highest vs. lowest quartile of serum B2: OR=0.76 (0.42-1.40), <i>p</i> for trend =0.68 highest vs. lowest quartile of serum B6: OR=1.33 (0.73-2.42), <i>p</i> for trend =0.43 highest vs. lowest quartile of serum folate: OR=0.67 (0.37-1.20), <i>p</i> for trend =0.19 highest vs. lowest quartile of serum B12: OR=0.87 (0.46-1.63), <i>p</i> for trend =0.98 highest vs. lowest quartile of serum homocysteine: OR=0.92 (0.51-1.66), <i>p</i> for trend = 0.60

RCC\*= renal cell carcinoma

<sup>a</sup> Multivariate-adjusted OR or RR and 95% confidence intervals (CIs) in parentheses

## Study designs in analytical epidemiology

Descriptive epidemiology concentrates on describing events (events or outcomes mostly refer to diseases) in population by characteristics related to time, place and individual<sup>213</sup>. From those observations, hypotheses are generated and tested through analytical approach. The principal objective of analytic epidemiology is to determine causes and effects, or the “why?” and the “how?” , through investigation of exposure and disease at the individual level<sup>214</sup>. A definition of this approach was further developed by Rothman *et al.*<sup>215</sup>, pointing that analytical epidemiology aims to evaluate whether particular exposure such as a diet or a biomarker, is associated or not to an event such as cancer, and then whether this association is independent. However, even if the association between the exposure and the event is statistically significant, causation is not necessarily present due to possible artefact, confounding, chance, or bias<sup>216</sup>. Nowadays, experimental and observational studies (mainly cohort and case-control studies) are the principal designs used in analytical epidemiology.

### Experimental studies

The specific characteristic of experimental studies is that the investigator controls the allocation of exposures to study subjects. Experimental studies, in particular clinical trials are superior to observational studies in the evidence hierarchy<sup>217</sup>. The reference of experimental studies is RCTs, where participants are randomly allocated to the groups (e.g. exposure or placebo). One advantage of such design is that potential unknown confounders are maintained symmetrically within the groups, and the comparability of the groups is maintained throughout the investigation. Thus, properly executed, RCT is the superior methodology to demonstrate causality and clinical efficacy of preventive and therapeutic procedures in the clinical setting.

Until now, the main focus of experimental studies of B-vitamins and cancer prevention has been on folate and colorectal cancer. Three RCTs of folate supplementation investigated whether it may prevent colorectal adenomas among high-risk population, but failed to identify a protective effect<sup>218-220</sup>. Although RCTs may restrict confounding, they have several limitations including; (1) limited size, limited number of cancers that occur in the follow-up period; (2) the exposure such as supplementation is randomized over a relatively short



period (usually shorter than few years); and (3) they are unrelated to lifelong nutrient levels prior to the study.

While the RCT may be considered as the “gold standard” of scientific studies, this design is not always suitable in epidemiological research, mainly because of ethical issues. It would for instance not be ethical to study the impact of a potential dangerous substance e.g. arsenic on kidney cancer by allocating arsenic to the exposed group and a placebo to the control group. Furthermore, RCTs overall are very costly and it may not be feasible to conduct them over many years.

In contrast to RCTs where investigators record outcome in relation to exposure that has been allocated in some random fashion, in observational studies, researchers limit themselves to observing the outcome in people who are already segregated into different exposure groups. The two main nonexperimental, observational study designs are cohort and case-control studies.

### **Cohort studies**

According to Breslow and Day<sup>221</sup>, *cohort studies are longitudinal studies and start by identifying a group of individuals about whom certain exposure information is collected; the group is then followed forward in time to ascertain the occurrence of the disease(s) of interest, so that for each individual prior exposure information can be related to subsequent disease experience.* Later, Rothman and Greenland<sup>222</sup> termed cohort as group (s) of subjects followed over time who share a common experience or status. For example, this could be a generation defined by the same period of birth, a group carrying a genetic trait (i.e. Down syndrome), a specific population who undergone special exposure (i.e. the Chernobyl cohort), or subjects gathered on a geographical basis (the Framingham cohort).

In public health, one the main objectives of cohort studies is to determine circumstances of disease occurrence and progression. First cohort investigations enrolled to assess the relation between risk factors and a disease of interest emerged during the 1950s in United-States<sup>223</sup> and in Grand Britain<sup>25 224</sup>. These studies were first devoted to evaluate the associations already strongly suggested by case-control studies or case-cohort studies. As an example, the British Doctors' Study from Doll and Hill was designed to assess the role of tobacco smoking on cancer mortality<sup>224</sup>. Usually, the cohort study design is referred to

exposed/unexposed studies because at least two groups of subjects (one exposed to a factor and the other not), without having initially the event, are followed and compared to the incidence of this event. The principle of this comparison is not far from experimental studies, although the exposition is not allocated in some random fashion by the investigators.

Once enrolled, all cohort members are classified according to their exposure status at the baseline (initial) examination, and these statuses can be updated during further examinations at different times throughout follow-up of participants.

Cohort participants can arise from different population sources. Cohort can originate entirely, or almost entirely, from a population defined geographically and temporally. For example, Hoffamns *et al.* carried out a cohort study” among the entire 1932 male birth cohort to investigate the impact of body mass index at the age of 18 on survival during the subsequent 32-year time interval<sup>225</sup>. However, it is more usual that the cohort is arising from a fraction or a sample of the population defined geographically and temporally. One well-known example concerns a cohort study that was initiated in the late 1940s, the Framingham Study, where participants were recruited in Framingham, Massachusetts. 6,507 adult subjects were selected to participate from a total list of 18,000 residents aged 35 to 59 years old in order to investigate risk factors of coronary heart disease<sup>223</sup>.

In order to correctly estimate the incidence of the disease(s) and its determinants, it is crucial to obtain accurate quality information regarding the disease occurrence. To ensure this quality information, document such as clinical records or death certificate of the participant are compared to previous documents recorded in interviews. In addition to validating the diagnosis of the disease, this comparison method permits to improve information regarding diagnosis characteristics, such as specific histological types in a cancer. A significant challenge in cohort studies is the validation of the cause of death, where either the physician is directly interviewed or information are collected from clinical records. Nevertheless, it is usually unfeasible to collect data on event for all members of the cohort due to some reasons including lost to follow-up (prior to the date of event) or death (from unrelated causes).

The design of a cohort study can vary and be prospective or retrospective. In a prospective cohort study, none of the participants have developed any of the event(s) of interest at study enrolment whereas in a retrospective cohort study (or historical study), exposure(s) and event (s) occurred before participants are enrolled in the study, and this specific cohort design can be defined for example from historic records.

In epidemiological methods, we can also differentiate closed cohorts from open cohorts. The main difference between these two types of cohorts is how membership is fixed. In a closed cohort, memberships are fixed and no participant can be added to the cohort once this latter started. Even if the size of the closed cohort diminishes due to death or loss to follow-up, no additional participant is further included. For example, people who were living in Hiroshima and Nagasaki when the atomic bombs were dropped in 1945 are part of a cohort whose membership started at the date of this nuclear disaster<sup>226</sup>, and no additional membership was included through time. These subjects remain in the cohort until they die. In contrast, open cohorts who dominate cancer epidemiology and form the conceptual basis for most case-control studies, are dynamic with members leaving or being added to the cohort. Rothman *et al.* has been given typical examples of open cohorts, in particular state cancer registries<sup>215</sup>. Every time a resident will be diagnosed with a cancer, he or she will be added to the cohort. Also, if the resident immigrates in another country or die, he or she can also leave the cohort.

In cohort studies, time is a key element and the concept of “person-time” is fundamental for design and analysis. Person-time corresponds to the number of study subjects and the time they are followed, or more explicitly the sum of all times contributed in a study by subjects at risk of a disease. Let’s assume for example that we have some participants exposed to arsenic water and were followed since they were exposed until the date of diagnosis, loss to follow-up, death or end of the study (censored date). Among those, 5 were followed 5 years and 10 were followed 2 years. This gives us a total person-time equal to 45 person-years ( $5*5 + 10*2 = 45$ ).

In cohort design, person-time is necessary to estimate disease risk between exposed and unexposed participants. The estimation can be relative that is about the degree by which the exposure increases risk (relative risk), or absolute that is about how much difference in

risk is there between the exposed and unexposed participants (attributable risk). The relative risk provides an idea of what is the strength of the association between the exposure and the disease which can guide for any potential causal relationship. Attributable risk indicates the burden of disease that an exposure displays in a population, which emphasize its public health impact<sup>213</sup>.

Many books and articles have reported on advantages and disadvantages of cohort studies<sup>215 222 227</sup>. In general, the main advantages are: exposures collected pre-diagnostically permit to avoid recall bias or reverse causality; they have the power to examine rare exposure and multiple events; incidence rates of disease in exposed and unexposed individuals can be determined. However, the cohort study design is not exempt of limitations such as (1) data on exposure and/or specimens for exposure analysis must be collected on the entire cohort (in prospective designs) at baseline which leads to cost issues and time consuming, (2) presence of bias from loss to follow-up (selection bias) between case and controls might induce the presence of an indirect non-causal association, and from outcome information being influenced by knowledge of exposures (information bias), (3) need to follow large number of subjects for long time to get enough statistical power, and (4) are limited for studies with rare events or long induction period.

### **Case-control studies**

Case-control studies are defined as those studies which compare the frequency of patients having the disease (cases) with subjects free of disease (controls) with regards to prior exposure(s) or characteristic(s)<sup>213</sup>. Case-control studies often constitute one of the first approaches to study the etiology of a disease or condition. The major objective of a case-control study is to strengthen hypothesized causal relationships by providing accurate and effective estimates.

Many epidemiological works reported on principal case-control methodologies<sup>213 215 222 227</sup> and include specification of the study hypothesis, definition and selection of cases and controls, collection of information, analysis and interpretation.

As with any other type of design, the study hypothesis as to be clearly defined earlier in order to avoid design weaknesses and problems when interpreting the results<sup>215</sup>.

A case-control study begins with the identification of cases which has to be ideally representative of all cases observed in a determined and identified population. Cases are selected on the basis of the disease and not the exposure. A case must be precisely and explicitly defined<sup>227</sup>. Objective evidence confirming the diagnosis of cases is usually advisable, even though the study must separate with some cases. For example, in cancer epidemiology, cancer cases are often defined thanks to the International Classification of Diseases (ICD) codes which confirm diagnosis of new incident cancer cases and limit ambiguity. Selection bias were raised by Rothman *et al.* when selecting all prevalent cases because those cases may not be representative of all cases<sup>215</sup>. For instance, the sub-population of prevalent cases with a short duration will be excluded from the entire group of prevalent cases because of either death or recovery and hence this might bias the relation between the exposure and the outcome. If documentation of incident cases is limited and non-precise this could lead to dilute the case group with some non-cases and the probability to discover real risk differences between cases and controls are diminished.

After cases identification, an important work has to be conducted for the selection of controls. Some epidemiologists noted that ideally controls should be directly sampled from the source population in order to have a well-representative selection. However, there are some debates pointing out that comparability of controls with cases might be more efficient. Rothman *et al.* discussed these two directions and argued that priority need to be done on the comparability of controls with the cases and that representativeness of the population is less important<sup>215</sup>. However, increasing attention on many levels is observed when considering population dimension. Conducting a population-based case-control study need generalisability, and to rich this, controls should be randomly selected from the disease free members of the same population from which all incident cases were drawn<sup>227</sup>. We could imagine for example all the newly diagnosed lung cancer cases in the catchment area of Lyon, assuming that they come from a regional cancer registry. Ideally, controls would suitably be drawn from the population of the same area within the same groups of age and sex than the cases. Sometimes, controls can also be drawn from hospitals although they are considered less suitable than population controls as they can lead to bias arising

from the factors which lead individual to use health care services. In addition to population and hospital, controls can be selected from other sources such as neighbours, friends, or relatives, although control selection should always be motivated by the assumption<sup>227</sup>.

An important aspect of case-control studies consists in determining the start and duration of exposure for cases and for controls. Several methodologies are used when recording data. Data on the relevant exposures can be obtained by telephone interview, by questionnaire, by face to face interview, by examining medical, occupational or other records, or by taking biological samples. Whatever method is chosen, it is crucial to guaranty that validated tools are used by trained investigators in the collection of data, and that such investigators should be "blind" (where possible) to the hypothesis under study and to the case/control status of the study participant. Even though this may be complex to realize in practice, strict procedures should be followed in order to limit bias from investigators (influence the cases and controls to remember exposures) and to maximized data collection.

Since the number of cases and controls is determined by the group conducting the research in traditional case-control studies, incidence rate ratio cannot be measured. However, if well-conducted, a case-control study can provide close estimates (odds ratio) of cohort studies (relative risk). The odds ratio is determined by the ratio of the odds of an exposure in the case group to the odds of an exposure in the control group<sup>215</sup>. Confidence interval for each odds ratio must be calculated in order to provide information on the precision of the estimates.

In addition, several statistical analysis methods are used in case-control studies in order to control for confounding or to evaluate interactions<sup>222</sup>. Measurement error and missing data are not rare in retrospective case-control studies and this can affect results. Sensitivity analysis which aims to quantitatively explore uncertainty in interpreting results has been proposed as one way of addressing these biases<sup>215 228</sup>.

Several methodological design books in epidemiology have been discussed limitations of case-controls studies<sup>213 215 222 227 229</sup>. When interpreting results, one should take into account potential biases including information bias in the data collection, selection bias in the selection of cases and controls, and oblivion of confounding factors<sup>216</sup>. For example, in

traditional case-control studies, cases will usually report the exposure information after learning of their diagnosis. Then, diagnosis may produce differential abilities in reporting, for example, by improving memory which might increase sensitivity among cases, or by inducing more false memory and this reduce sensitivity. These characteristics are examples of recall bias. However, recall bias can be limited by keeping the study participants unaware of the hypotheses under study.

If cases and controls are not representative of the source population from where cases arise, this could induce distortion in the results (selection bias). Despite approaches in selecting population-based controls which limit this type of bias, clear and precise monitoring need to be performed on response rates and participation bias<sup>227</sup>.

Another concern in case-control studies is that diagnosis related behavioural or pre-malignant changes may affect the exposure, such as biomarkers close to diagnosis, i.e. as would be expected by reverse causality<sup>230</sup>.

One of the major issue when conducting analytical epidemiology and in particular case-control study is question related to validity and precision. First, several methods can help in improving results precision or diminishing random errors by obtaining large sample size population (using pre-study power analysis) or by improving study design (for example matching control group with cases with respect to certain characteristics other than the exposure under study such as age or sex). In addition, one could think to validate results internally by conducting several sensitivity analysis (accuracy of measurement), and also to validate them externally by conducting the same study outside that population (generalisability of findings)<sup>231</sup>.

In addition, confounding is a considerable important issue in nonexperimental research. On the simplest level, confounding is a confusion of effects of the exposure of interest<sup>231</sup>. The distortion introduced by a confounding factor can lead to over- or underestimation of an association, or consequently more, can change the direction of the association. It is therefore essential to control for confounding throughout each investigation notably in observational study designs compared to intervention studies.

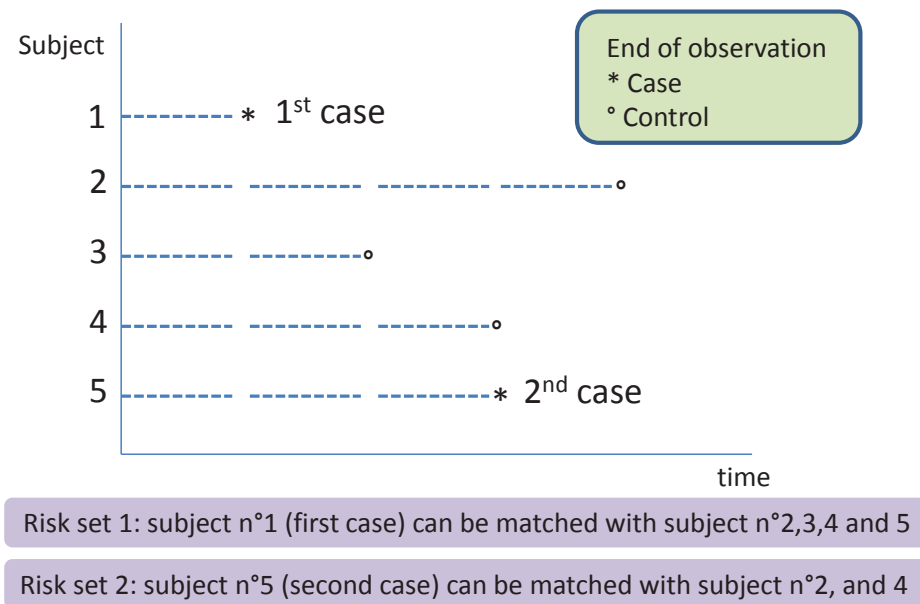
Although traditional case-control designs are more likely to have bias and incorrect inference compared to cohort studies, several epidemiologists highlighted important

advantages of case-control designs including that they are cost-effective, easy to conduct, relatively quick, very suitable for rare events (e.g. cancer), and there is no participant that can be lost to follow-up<sup>215 227 230</sup>. If well-designed and well-conducted, case-control studies permit to evaluate multiple hypotheses and interactions which can make this design efficient.

### **Nested case-control studies**

We explained previously that in traditional case-control studies, controls should be selected to be representative of the study population source from which the cases were drawn. If they are not, selection bias will be introduced. A variant of case-control studies, named “nested case-control studies”, permit readily to accomplish this selection. As the name implies, a nested case-control study is a case-control study nested within a cohort, in which the cases and the controls arise from a clearly defined population, the cohort. However, in order to be able to adapt such design, it is important to enumerate times of entry and exit known for all members in order to list each risk set. These risk sets are necessary to ensure a rough representativeness when using sampling methods. In general, the selection method mostly used in nested case-control study is the incidence density sampling<sup>229</sup>. This method aims to sample controls from the risk set at the time each case is diagnosed and that this risk set is completely independent from the other risk sets. In order to improve comparability between cases and controls, they might be matched on various demographic characteristics such as sex, age, or year of recruitment. Importantly, a control being sampled from the population at time  $t$  could become afterwards a case at a time  $t+n$ , and also the same participant could be sampled more than once as a control. [Figure 11](#) shows a simple situation of two risk sets. For the first case (risk set n°1) 4 controls are eligible for the matching and for the second case (risk set n°2) only 2 controls remained eligible. Similar density sampling methods were applied throughout the different nested case-control studies used in this dissertation.





**Figure 11.** Representation of risk sets in nested case-control designs.

When using this approach for case-control matching, the association estimate (odds ratio) from the nested case-control design is an unbiased estimate of constant incidence rate ratio from the cohort<sup>232</sup>. As an example, Hak E *et al.* and colleagues compared relative and absolute estimates of associations in an influenza vaccine study using both cohort and nested case-control designs<sup>233</sup>. After selection of different ratios (1:1 to 1:4) of cases and matched controls, it has been observed that the nested case-control design provided valid and precise estimates of associations compared to the full-cohort analysis.

Because the nested design conserves the validity of the prospective study, bias due to the difference in recall and bias of the disease modifying biological characteristics as mentioned for traditional case-control designs are avoided. Also, nested case-control studies compared to cohort design are more efficient in a way that they can give similar statistical and significant estimation with a smaller sample size than required for analysis in cohorts. By selecting a sample of control(s) for each case, the number of participants for whom exposure information needs to be collected, such as a biomarker, is reduced, which might decrease cost and logistics compared to obtain the exposure of interest in the entire cohort<sup>234</sup>. However, important data including baseline characteristics must be collected at enrolment for all participants in the cohort making the costs of nested design more likely higher than a traditional case-control study.

In line with the advantages of the nested design, in particular time and cost, we conducted several nested case-controls studies in order to contribute in understanding the role of OCM biomarkers in relation to smoking-related cancers.

# **CHAPTER II:**

## ***MATERIAL AND METHODS***

## Data sources

### The European Prospective Investigation into nutrition and Cancer (EPIC)

#### *Overview of cohort*

In 1992, the International Agency for Research on Cancer (IARC) initiated a comprehensive collaboration across Europe, the EPIC cohort, a multi-centre prospective cohort study in Western Europe<sup>12 235</sup>. The main objective of the study was to investigate the etiology of human cancer of various sites in relation to nutritional and other lifestyle factors, further motivated by difficulties in measuring dietary exposures and the need for very large prospective studies in order to establish associations between these factors and disease. Originally, seven EPIC countries were included in the project including France, Germany, Greece, Italy, The Netherlands, Spain and the United Kingdom. These core countries were subsequently followed by three Nordic countries, Denmark, Sweden and Norway comprising in total 10 countries and 23 individual centres across Europe (see [Figure 12](#)).



**Figure 12.** EPIC study collaborating centres.

The enrolment of participants from all EPIC centres took place between 1992 and 2000. In general, women and men aged 35 to 70 years were invited to participate either by mail or in

person. In total, the EPIC study has recruited 521,330 participants from all participating centres.

#### *Study population and data collection*

Individuals who agreed to participate signed an informed agreement and received two questionnaires on dietary and lifestyle variables, respectively. As a whole, subjects completed questionnaires at home and were then invited to a study centres for additional examination including anthropometric measurements and blood samples donation. From all participants recruited in the EPIC cohort, 385,747 provided blood samples, which were collected to a standardised protocol. After the blood draw, fractions of blood (plasma, serum, erythrocytes and buffy coat for DNA) were aliquoted into several 0.5 mL straws. One set containing half of the straws was stored locally and the remaining straws were transported to IARC biobank (Lyon) for storage in liquid nitrogen (-196°C). However, with exception of the Danish and Umeå (Sweden) samples that were stored in plastic tubes locally. The Danish samples were stored in liquid nitrogen tanks in the biobank of The Danish Cancer Society in Copenhagen, and the samples from the Umeå study centre were stored in -80°C freezers in Umeå.

Different methods of dietary assessment were adopted in the EPIC cohort. First, extensive self-administrated quantitative dietary questionnaires, containing more than 250 food items and estimating individual average portions systematically, were used in several countries including northern Italy, The Netherlands, Germany and Greece (where dietary questionnaires were interviewer-administered). Questionnaires, similar in content to the self-administered quantitative dietary questionnaires but structured by meals, were used in Spain, France and Ragusa (south Italy). To increase compliance, the centres in Spain and Ragusa performed a face-to-face dietary interview using a computerised dietary program, whereas the dietary questionnaire was self-reported in France.

Then, semi-quantitative food-frequency questionnaires (with the same standard portion(s) assigned to all subjects) were used in Denmark, Norway, Naples in Italy and Umeå<sup>o</sup> in Sweden. Finally, combined dietary methods were used in the UK and Malmö (Sweden). The two British centres used both a semi-quantitative food-frequency questionnaire and a 7-day record, whereas a method combining a short non-quantitative food-frequency

questionnaire with a 14-day record on hot meals (lunches and dinners) was developed in Malmö.

The ascertainment of cases within EPIC reflects the predominant pattern of diagnosis in the general population because the majority of cohort's subjects were invited from the general adult population residing in a given town or geographical area. Incident cancer cases were identified at regular intervals through population-based cancer registries (Denmark, Italy except Naples, the Netherlands, Norway, Spain, Sweden and United Kingdom) or by active follow-up (France, Germany, Greece and Naples), which involved a combination of methods, including review of health insurance records, cancer and pathology registries, as well as direct contact with participants and their next-of-kin.

Mortality data, including vital status, cause of death, and date of death, were obtained from mortality registries at the regional or national level. Subjects were followed up from study entry until cancer diagnosis (except non-melanoma skin cancer), death, emigration, or the end of the follow-up period for the relevant study centre. End of follow-up was defined as the latest date of complete follow-up for both cancer incidence and vital status and varied between study centres from December 2004 to June 2010. Vital status at follow-up is over 98% complete.

### **The Lung Cancer Cohort Consortium (LC3)**

#### *Overview of cohorts*

The LC3 was initially created in 2011. This initiative followed an investigation within the EPIC study where it was reported strong associations between circulating concentrations of vitamin B6, folate and methionine and lung cancer risk<sup>13</sup>. The overarching objectives of this large consortium are two-fold; (1) to evaluate potential etiological risk factors of lung cancer over and above smoking history, and (2) to develop extensive risk prediction models using biomarkers in multiple cohorts. The LC3 was initiated in the collaboration of several cohorts participating in the National Cancer Institute (NCI). All cohorts with at least 150 lung cancer cases having prospectively collected questionnaire data and either plasma or serum samples were invited to participate to the consortium. Many cohorts fulfilled those criteria and accepted to participate, and to date, the LC3 includes 20 prospective cohorts with a

combined study population of over 2,000,000 participants from North America, Europe, Asia and Australia, and are listed in the table below. Further design information and follow-up procedures of each cohort are provided in Appendix 1.

**Table 5.** Characteristics of participating cohorts and lung cancer cases with serum/plasma available

Nb.	Full Cohort name	Abbr.	Representative (co-applicant)	Location	Enrolment	Cohort size
Cohorts recruited in US						
1	Women’s Health Initiative	WHI	Ross Prentice	Fred Hutchinson CRC	1994-1998	161,808
2	Southern Community Cohort Study	SCCS	Qiuyin Cai Shiman Qu	Vanderbilt University Medical Center	2002-2009	86,000
3	Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial	PLCO	Neil Caporaso	DCP/DCEG, NCI, US	1993-2001	155,000
4	NY Univ. Women Health Study	NYUWHS	Alan Arslan Anne Zeleniuch-Jacquotte	NYU School of Medicine	1985-1991	14,274
5	ACS Cancer Prevention Study-II	CPS-II	Victoria Stavens Marji McCullough	American Cancer Society	1992	184,190
6	Clue studies 1974 CLUE (CLUE I) & 1989 (CLUE II)	CLUE	Kala Visvanathan	John Hopkins Bloomberg School of Public Health	1974 & 1989	50,000
7	Multiethnic Cohort Study	MEC	Loic le Marchand Lynne Wilkens	Cancer Research Center, University of Hawaii	2001-2007	215,000
8	Women’s Health Study	WHS	Shuming Zhang	Brigham and Women's Hospital / Harvard Uni	1992-1995	39,876
9	Nurses' Health Study	NHS I	Jiali Han Xuehong Zhang	Brigham and Women's Hospital / Harvard Uni	1989	32,826
10	Health Professionals Follow-up Study	HPFS	Jiali Han	Harvard University	1993	18,159
11	Physicians health study I & II	PHS I+II	Jing Ma Howard Sesso	Brigham and Women's Hospital	1982-1983 1997-2000	29,071
Cohorts recruited in Europe/ Australia						
12	Melbourne Collaborative Cohort Study	MCCS	Gianluca Severi Graham Gilles	Cancer Epidemiology Centre / Victoria	1990-1994	41,514
14	Malmö Diet and Cancer	MDCS	Jonas Manjer	Lund University,	1991-1996	28,098

study			Sweden			
15	Northern Sweden Health and Disease Cohort	NSHDC	Mikael Johansson Kjell Grankvist	Umeå University, Sweden	1985-onward	99,100
16	Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study	ATBC	Demetrius Albanes Stephanie Weinstein	DCEG, NCI, US	1985	29,133
17	The HUNT study	HUNT	Arnulf Langhammer Kristian Hveem	HUNT Biosciences, Norway	1995-1997	100,00
Cohorts recruited in Asia						
18	Shanghai Cohort Study	SCS	Jian-Min Yuan	University of Minnesota	1986-1989	18,244
19	Singapore Chinese Health Study	SCHS	Lesley M. Butler	University of Minnesota	1993-1998	63,257
20	Shanghai Men's Health Study	SMHS	Xiao-Ou Shu	Vanderbilt University	2001-2006	61,500
21	Shanghai Women's Health Study	SWHS	Wei Zheng	Vanderbilt University	1996-2000	74,942

### *Study population and data collection*

In total, the consortium included 11,399 incident lung cancer cases with available pre-diagnostic blood samples. Of these, we selected a subsample of 5,545 cases, comprising all incident cases among never smoking participants (n=1336), and a random selection of the cases diagnosed among former and current smoking participants. A standardised data base dictionary was sent to each cohort and we requested a standardized set of covariates for each study subject, including variables on (1) anthropometric and demographic variables such as weight, height, ethnicity, smoking status, educational attainment; (2) dietary intake variables such as vegetables, fruit, vitamins supplements; (3) clinical variables such as tumour stage, histology, date of diagnosis, and (4) follow-up information such as date of death and cause of death. All plasma and serum samples were sent and centralized to the Bevitall AS laboratory in Bergen, Norway, and once on site they were kept at -80°C. A minimum of 500 µL of blood fraction (plasma or serum) from all participating subjects in each individual cohort was requested.

### **Biochemical analyses**

All serum/plasma samples from the studies were shipped and centralized to Bergen under appropriate conditions, using established protocols (e.g sent on dry ice). Sample volume



requirements have been minimized by using a combination of microbiological assays and mass-spectrometry based methods, and sample handling was carried out by robotic workstations ensuring high sample throughput. Measurements performed at Bevital have been used in several high profile, recently published epidemiological studies<sup>13 236-238</sup>. All biochemical analyses of circulating OCM biomarkers were performed at the the Bevital A/S laboratory (<http://www.bevital.no>) in Bergen, Norway. Plasma from all study participants were analyzed on four analytical platforms (platforms B, C, D, and F). Concentration measurements of plasma vitamin B2 (riboflavin), vitamin B6 (measured as pyridoxal 5'-phosphate [PLP], its active form), vitamin B9 (folate), vitamin B12 (cobalamin), total homocysteine, and methionine were provided in HNC (Chapter III) and RCC (Chapter IV) studies. LC3 study (Chapter V) comprised plasma/serum measurement concentrations of vitamin B6, folate and methionine. All case and control participants were successfully analysed, for at least one of the biomarker. Circulating cotinine was also measured as an indicator of recent smoking behaviour in each study. Circulating vitamin B2, vitamin B6, homocysteine, methionine, and cotinine were determined by mass spectrometry-based methods (liquid chromatography coupled to tandem mass spectrometry; gas chromatography coupled to tandem mass spectrometry - LC-MS/MS, GC-MS/MS)<sup>239 240</sup>. Microbiological methods were used to determine concentrations of folate (*Lactobacillus casei*)<sup>241</sup> and vitamin B12 (*Lactobacillus leichmannii*)<sup>242</sup>.

The detection limit, and within-day and between-day coefficients of variations (Cvs) using these methods are low for all of the biomarkers<sup>239 240</sup>. For HNC and RCC studies, CVs within and between batches were, respectively, 6 and 11% for vitamin B2, 3% and 6% for vitamin B6, 4% and 5% for folate and vitamin B12, 1% and 3% for methionine, and 1% and 2% for homocysteine. The CVs within and between batches of biomarkers studied in LC3 study were, respectively, 3 and 7% for vitamin B6, 4 and 5% for folate and 1 and 3% for methionine.

All samples were analysed with similar methods in batches of 86 and quality control included six calibration samples, two control samples, and one blank sample in each batch. For each study, all serum or plasma samples were kept at -80°C and all cancer cases and their individual matched controls were analysed together within the same batches in random order. The laboratory staff was blinded to the case-control status of the blood samples. The raw data from the separate platforms and analysis sets were merged, handled

and controlled by specifically designed software. Outliers were identified by the ratio of concentrations as measured by the different platforms. Unlikely combinations of values were flagged by installed macros. This increases accuracy and precision, minimizes the possibility of assay interference and ensures adequate sample handling and logistics. The system also generated summary statistics and printouts of selected samples. These routines for data handling ensured efficiency and reduced the chance of logistic problems and human errors.

In the LC3 study, about 1,000 repeat blood samples from subjects were additionally measured in order to estimate the within person variation in each biomarker. These repeat samples came from two participating cohorts which are the CSP (390 repeat samples) and NSHDS (600 repeat samples) studies. The sample size of repeat samples have been designed to select approximately one third of repeat samples collected within a short time of each other (< 1 year), as well as one third collected around 3 to 5 years of each other, and one third within 10 years or more of each other.

## **Statistical methods**

Standard statistical methods, such as linear and logistic regression, were used throughout the different studies of this thesis in order to investigate various relationships. The strength of associations was estimated by OR. P-values were assessed as indicators of associations, and p-values of less than 0.05 were considered statistically significant.

Throughout the papers included in this thesis, we conducted initial analyses involving extensive assessments of demographic variables such as age, smoking status, body mass index (BMI) and factor related to socio economical status as well as for dietary intake of fruit and vegetables. Also, the relation of lifestyle and dietary factors with biomarker concentrations were assessed using linear models, adjusted for age, sex, country and other covariates if necessary. Subsequent statistical analysis included estimation of RR by calculating OR, using both conditional and unconditional logistic regression. In the three studies of this thesis, risk analysis involved calculating quartiles of circulating concentration for each biomarker of interest based on their distribution among all matched control

participants. Because of the substantial variation of circulating biomarkers across continent in LC3 study (Chapter V), quartiles were additionally based on the distribution among all controls of each continent (United States [US], Europe and Australia combined [EU/AU], and Asia). OR and 95% CI for participants in the second, third, and fourth quartiles were calculated relative to the first quartile using conditional logistic regression, conditioning on individual case sets. Unconditional logistic regression were also used mostly in HNC and RCC studies for stratified analysis or in order to increase the statistical power (since we had about half controls who were not matched to the cases, see Selection of cases and controls in Chapter III and IV) and further evaluate the consistency of any association observed in conditional logistic regression.

In order to evaluate whether deficiency of the OCM biomarkers of interest in LC3 study are harmful for lung cancer etiology, we conducted analyses by using clinical cut-points for deficiency of vitamin B6 and folate. Because methionine deficiency is very poorly documented, we could not conduct such analysis. We have considered vitamin B6 deficiency as circulating vitamin B6 concentration less than 20nmol/L, compared with normal range, defined as  $\geq 20\text{nmol/L}$ <sup>243</sup>. Folate was classified as deficient ( $<7\text{ nmol/L}$ ), moderate deficiency (7 to 13 nmol/L), or normal ( $\geq 13\text{ nmol/L}$ )<sup>244</sup>. For these additional analyses in the LC3 study, we also used conditional logistic regression to estimate RRs by calculating ORs and 95% CI of lung cancer risk for deficiency relative to normal concentrations.

To assess potential effect modification throughout the studies, we performed stratified analyses by different variables including place of recruitment (country or continent), sex, smoking status, age at diagnosis, time from blood draw to diagnosis.

In order to investigate if OCM biomarkers were related to cancer survival probability of the head and neck and kidney, estimation of hazard ratios (HRs) for all-cause mortality for cases were calculated using Cox proportional hazards models (Chapter III and IV, respectively). Time since diagnosis was used as the timescale, and all models were adjusted for age at diagnosis, sex, and country and additional variables when it was relevant.

We adjusted overall analysis by cotinine (in quartiles) within each study, by smoking status (in HNC and RCC studies), and further smoking variables when available, including duration of smoking and average cigarettes smoked per day. Further potential confounders, including

alcohol consumption, BMI, educational attainment, hypertension and dietary intake were included in our analyses where necessary.

For estimated interclass correlations of repeat measurements in the LC3 study, we used spearman correlation methods<sup>215</sup>.

All statistical analyses were conducted using SAS 9.2 (Cary, NC)<sup>245</sup> (Chapter III, IV and V), Stata 12.1 for Linux (Stata Corporation)<sup>246</sup> (Chapter IV), R version 3.1.3<sup>247</sup> (Chapter V), and Stan version 2.6.0<sup>248</sup> (Chapter V).

# **CHAPTER III:**

## ***ONE-CARBON METABOLISM AND HEAD AND NECK CANCER***

## Objective

Although our group has recently reported strong inverse associations between circulating concentrations of OCM biomarkers, in particular vitamin B6, folate and methionine, with risk of lung cancer, the risk association of OCM biomarkers in other smoking related-cancers remains to be demonstrated.

In this Chapter we primarily aim to investigate whether pre-diagnostic circulating concentrations of vitamin B2, vitamin B6, folate, vitamin B12, methionine and homocysteine are associated with onset and survival of HNC. Additionally, we investigated the relation of those OCM factors with ESCC risk and mortality.

## Selection of cases and controls

This study was nested in the EPIC cohort and further study design information has been described previously in Chapter II (see *Material and Methods, Study population and data collection*).

In this study, 1,273 subjects diagnosed with incident head and neck or oesophagus cancers were identified from nine EPIC sub-cohorts contributing to this nested case-control study. These cancer cases were defined on the basis of the International Classification of Diseases for Oncology, Second Edition (ICD-O-2), and included: oral cavity (ICD C02.0-C02.9, C04.0-C04.9, C03.0-C03.9, C05.0-C06.9, C14.0-C14.9), oropharynx (C01.9, C02.4, C09.0-C10.9), hypopharynx (C13.0- C13.9), larynx (C32.0-C32.9), and oesophagus (C15.0-C15.9). Criteria of exclusion included cases who (1) did not donate a blood sample (n=152), (2) did not have enough plasma available for biochemical analysis (n=20), (3) had a history of another cancer (n=158, except non melanoma skin cancer), (4) were not histologically confirmed, were prevalent at the time of blood donation, or did not have questionnaire information available (n=22), (5) cases from Denmark (n=288) because of flooding of the Danish biobank, and the Malmo centre in Sweden (n=101) who did not participate in this study. We thought also to exclude adenocarcinoma of the head and neck (n=16) because the etiology of adenocarcinoma and SCC are likely to differ and the vast majority of the HNC are SCC. After all participant exclusions, 516 cancer cases of the head and neck and oesophagus remained eligible in this study.

For each case, one control was randomly chosen from appropriate risk sets consisting of all cohort members alive and free of cancer (except non-melanoma skin cancer) at the time (and hence age) of diagnosis of the index case. Matching criteria were country, sex, date of blood collection ( $\pm 1$  month, relaxed to  $\pm 5$  months for sets without available controls), and date of birth ( $\pm 1$  year, relaxed to  $\pm 5$  years for sets without available control participants). An additional control group (control group 2, n=479) was included in the HNC study. This control group 2 was matched to cases of the study on RCC (from Chapter IV) using similar matching criteria. Control group 2 participated essentially to unconditional and stratified analyses in order to assess the consistency of any association and increase the statistical power.

### **Biochemical analyses**

All biomarkers were sent on dry ice to the Bevital A/S laboratory, in Bergen, Norway. Because the studies of this thesis share common biochemical methods, details of the biochemical analyses have been provided previously in Chapter II (see *Material and Methods, Biochemical analyses*).

### **Statistical analyses**

Throughout this study, we used both conditional and unconditional logistic regression in order to investigate circulating OCM biomarkers in relation to cancer risk of the HNC and ESCC. Hazard ratios for all-cause mortality for HNC and ESCC cases were calculated using Cox proportional hazard regression models. These calculations were performed using SAS 9.2. More details in statistical methods are provided in Chapter II (see *Material and Methods, Statistical methods*).

# Article 1

## A prospective study of one-carbon metabolism and cancer of the head and neck and oesophagus.

**Anouar Fanidi**, Caroline Relton, Per Magne Ueland, Øivind Midttun, Stein Emil Vollset, Ruth C. Travis, Antonia Trichopoulou, Pagona Lagiou, Dimitrios Trichopoulos, H.B(as). Bueno-de-Mesquita, Martine Ros, Heiner Boeing, Rosario Tumino, Salvatore Panico, Domenico Palli, Sabina Sieri, Paolo Vineis, María-José Sánchez, José María Huerta, Aurelio Barricarte Gurrea, Leila Luján-Barroso, J Ramón Quirós, Anne Tjønneland, Jytte Halkjær, Marie-Christine Boutron-Ruault, Françoise Clavel-Chapelon, Claire Cadeau, Elisabete Weiderpass Vainio, Mikael Johansson, Elio Riboli, Paul Brennan, Mattias Johansson

*International Journal of Cancer* 2015 Feb 15; 136(4):915-27





IJC

International Journal of Cancer

## A prospective study of one-carbon metabolism biomarkers and cancer of the head and neck and esophagus

Anouar Fanidi<sup>1</sup>, Caroline Relton<sup>2,3</sup>, Per Magne Ueland<sup>4,5</sup>, Øivind Midttun<sup>6</sup>, Stein Emil Vollset<sup>7,8</sup>, Ruth C. Travis<sup>9</sup>, Antonia Trichopoulos<sup>10,11</sup>, Pagona Lagiou<sup>10,11,12</sup>, Dimitrios Trichopoulos<sup>11,12,13</sup>, H.B(as). Bueno-de-Mesquita<sup>14,15</sup>, Martine Ros<sup>14</sup>, Heiner Boeing<sup>15</sup>, Rosario Tumino<sup>16</sup>, Salvatore Panico<sup>17</sup>, Domenico Palli<sup>18</sup>, Sabina Sieri<sup>19</sup>, Paolo Vineis<sup>20,21</sup>, María-José Sánchez<sup>22,23</sup>, José María Huerta<sup>23,24</sup>, Aurelio Barricarte Gurrea<sup>25,26</sup>, Leila Luján-Barroso<sup>27</sup>, J Ramón Quirós<sup>28</sup>, Anne Tjønneland<sup>29</sup>, Jytte Halkjær<sup>29</sup>, Marie-Christine Boutron-Ruault<sup>30</sup>, Françoise Clavel-Chapelon<sup>31</sup>, Claire Cadeau<sup>32</sup>, Elisabete Weiderpass<sup>33</sup>, Mikael Johansson<sup>34</sup>, Elio Riboli<sup>20</sup>, Paul Brennan<sup>1</sup> and Mattias Johansson<sup>1</sup>

<sup>1</sup> International Agency for Research on Cancer, Lyon, France

<sup>2</sup> Institute of Genetic Medicine, Newcastle University, Newcastle, United Kingdom

<sup>3</sup> MRC Integrative Epidemiology Unit, School of Social & Community Medicine, University of Bristol, Bristol, United Kingdom

<sup>4</sup> Section of Pharmacology, Institute of Medicine, University of Bergen, Bergen, Norway

<sup>5</sup> Laboratory of Clinical Biochemistry, Haukeland University Hospital, Bergen, Norway

<sup>6</sup> Bevitall AS, Bergen, Norway

<sup>7</sup> Department of Public Health and Primary Health Care, University of Bergen, Bergen, Norway

<sup>8</sup> Division of Epidemiology, Norwegian Institute of Public Health, Bergen, Norway

<sup>9</sup> Nuffield Department of Clinical Medicine, Cancer Epidemiology Unit, University of Oxford, Oxford, United Kingdom

<sup>10</sup> Department of Hygiene, Epidemiology and Medical Statistics, WHO Collaborating Center for Food and Nutrition Policies, University of Athens Medical School, Athens, Greece

<sup>11</sup> Department of Epidemiology, Harvard School of Public Health, Boston, MA

<sup>12</sup> Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece

<sup>13</sup> Hellenic Health Foundation, Athens, Greece

<sup>14</sup> National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

<sup>15</sup> Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany

<sup>16</sup> Cancer Registry and Histopathology Unit, "Civile M.P. Arezzo" Hospital, ASP Ragusa, Italy

<sup>17</sup> Dipartimento di medicina clinica e chirurgia, Federico II University, Naples, Italy

<sup>18</sup> Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute—ISPO, Florence, Italy

<sup>19</sup> Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

<sup>20</sup> School of Public Health, Imperial College London, London, United Kingdom

<sup>21</sup> HuGeF Foundation, Turin, Italy

<sup>22</sup> Andalusian School of Public Health, Granada, Spain

<sup>23</sup> Consortium for Biomedical Research in Epidemiology and Public Health (CIBER Epidemiología y Salud Pública-CIBERESP), Madrid, Spain

<sup>24</sup> Department of Epidemiology, Murcia Regional Health Council, Murcia, Spain

**Key words:** one-carbon metabolism, head and neck cancer, EPIC

**Abbreviations:** CI: confidence interval; CRC: colorectal cancer; CV: coefficient of variations; EPIC: European Prospective Investigation into Cancer and Nutrition; ESCC: esophagus squamous cell carcinoma; HNC: head and neck cancer; HR: hazard ratio; OCM: one-carbon metabolism; OR: odds ratio; SCC: squamous cell carcinoma

Additional Supporting Information may be found in the online version of this article.

**Grant sponsors:** World Cancer Research Fund (UK), Europe Against Cancer Program of the European Commission (SANCO), Deutsche Krebshilfe; Deutsches Krebsforschungszentrum, German Federal Ministry of Education and Research, Danish Cancer Society, Health Research Fund (FIS) of the Spanish Ministry of Health, Spanish Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, Catalan Institute of Oncology, Spain and the ISCIII of the Spanish Ministry of Health; **Grant number:** RETICC DR06/0020; **Grant sponsors:** Cancer Research UK, Medical Research Council, United Kingdom, Greek Ministry of Health, Stavros Niarchos Foundation, Hellenic Health Foundation, Italian Association for Research on Cancer (AIRC), Italian National Research Council, Fondazione-Istituto Banco Napoli, Italy, Associazione Italiana per la Ricerca sul Cancro-AIRC-Milan, Compagnia di San Paolo, Dutch Ministry of Public Health, Welfare and Sports, World Cancer Research Fund, Swedish Cancer Society, Swedish Scientific Council; **Grant sponsor:** Regional Government of Västerbotten, Sweden, NordForsk (Centre of excellence programme HELGA), Norway, French League against Cancer (LNCC), France, National Institute for Health and Medical Research (INSERM), France, Mutuelle Générale de l'Éducation Nationale (MGEN), France, 3M Co., France, Gustave Roussy Institute (IGR), France, General Councils of France

DOI: 10.1002/ijc.29051

History: Received 29 Jan 2014; Accepted 20 May 2014; Online 28 June 2014

Correspondence to: Mattias Johansson, Genetic Epidemiology Group, International Agency for Research on Cancer (IARC/WHO), 150 cours Albert Thomas, 69008, Lyon, France, Tel.: +33-4-72-73-80-23, Fax: +33-4-72-73-83-42, E-mail: JohanssonM@iarc.fr

Int. J. Cancer: 00, 00–00 (2014) © 2014 UICC

<sup>25</sup> Consortium for Biomedical Research in Epidemiology and Public Health (CIBER Epidemiología y Salud Pública-CIBERESP), Barcelona, Spain

<sup>26</sup> Navarra Public Health Institute, Pamplona, Spain

<sup>27</sup> Institute of Oncology (ICO-IDIBELL), Barcelona, Spain

<sup>28</sup> Public Health Directorate Asturias, Oviedo, Spain

<sup>29</sup> Danish Cancer Society Research Center, Copenhagen, Denmark

<sup>30</sup> Inserm U1018, Centre for research in Epidemiology and Population Health (CESP), Villejuif, France

<sup>31</sup> University Paris Sud, UMR5 1018, Villejuif, France

<sup>32</sup> Institut Gustave Roussy (IGR), Villejuif, France

<sup>33</sup> Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, Tromsø, Norway

<sup>34</sup> Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden

Experimental and epidemiological data suggest that factors of one-carbon metabolism are important in the pathogenesis of several cancers, but prospective data on head and neck cancer (HNC) and esophagus cancer are limited. The European Prospective Investigation into Cancer and Nutrition (EPIC) study recruited 385,747 participants from 10 countries who donated a blood sample. The current study included 516 cancer cases of the head and neck and esophagus and 516 individually matched controls. Plasma levels of vitamins B2, B6, B9 (folate), B12, and methionine and homocysteine were measured in pre-diagnostic plasma samples and analyzed in relation to HNC and esophagus cancer risk, as well as post-diagnosis all-cause mortality. After controlling for risk factors, study participants with higher levels of homocysteine had elevated risk of HNC, the odds ratio (OR) in conditional analysis when comparing the top and bottom quartiles of homocysteine [OR<sub>Q4 vs. Q1</sub>] being 2.13 (95% confidence interval [95% CI] 1.13–4.00, *p* for trend 0.009). A slight decrease in HNC risk was also seen among subjects with higher levels of folate (OR<sub>Q4 vs. Q1</sub> 0.63, 95% CI 0.35–1.16, *p* for trend 0.02). Subgroup analyses by anatomical sub-site indicated particularly strong associations with circulating homocysteine for oral cavity and gum cancer (*p* for trend  $8 \times 10^{-4}$ ), as well as for oropharynx cancer (*p* for trend 0.008). Plasma concentrations of the other investigated biomarkers did not display any clear association with risk or survival. In conclusion, study participants with elevated circulating levels of homocysteine had increased risk of developing squamous cell carcinoma of the head and neck.

#### What's new?

One-carbon metabolism (OCM) involves the transfer of a carbon unit from methyl donor nutrients to molecules involved in the synthesis and methylation of DNA. As a result, dietary imbalances or deficiencies in nutrients crucial for OCM may affect DNA replication, repair, and regulation, potentially facilitating cancer development. This analysis of circulating levels of OCM nutrients in head and neck cancer and esophageal cancer patients and matched controls reveals an association between elevated levels of the amino acid homocysteine and increased risk of squamous cell carcinoma of the head and neck. Risk was decreased slightly by elevated folate levels.

In 1981 Doll and Peto estimated that approximately 35% of cancer deaths in the United States could be avoided by modification of diet.<sup>1</sup> While little evidence has subsequently emerged on food groups linked to specific cancer sites, cancers of the head and neck (HNC) consistently occur less frequently among subjects with a high consumption of fruits and vegetables, with supporting data from both case-control<sup>2</sup> and prospective studies.<sup>3,4</sup>

Fruits and vegetables are important dietary sources of some B-vitamins and additional nutrients that are involved in the one-carbon metabolism (OCM).<sup>5–7</sup> The metabolic pathway of OCM has been frequently implicated in carcinogenesis because of its involvement in maintaining nucleotide synthesis and methylation. Imbalances and deficiencies among crucial OCM nutrients may interfere with DNA repli-

cation, DNA repair and regulation of gene expression, any of which could promote carcinogenesis.<sup>8,9</sup>

The one-carbon metabolism pathway has been frequently investigated in relation to multiple cancer types in prospective studies,<sup>10–14</sup> but not yet in relation to head and neck cancer. Recently, Johansson *et al.* reported that elevated serum levels of both vitamin B6 and methionine were associated with a reduced risk of lung cancer of approximately 50%, independently of smoking status.<sup>12</sup> These results, as well as similar studies on other cancer sites such as colorectal cancer,<sup>15</sup> suggest that the one-carbon metabolism biomarkers may provide important information on disease risk for some specific cancers.

In order to obtain evidence on the importance of one-carbon metabolism in HNC and esophagus cancer, we

conducted a large nested case-control study within the European Prospective Investigation into Cancer and Nutrition (EPIC).

### Material and Methods

#### Study cohort

The study was nested within the European Prospective Investigation into Cancer and Nutrition (EPIC) study. EPIC recruitment procedures, collection of questionnaire data, anthropometric measurements, and blood samples have been described in detail.<sup>16</sup> In brief, 521,330 individuals were recruited to the cohort between 1992 and 2000 from 10 European countries, of whom 385,747 contributed a blood sample. Blood fractions were aliquoted into 0.5 mL straws, which were heat-sealed and stored in liquid nitrogen tanks at  $-196^{\circ}\text{C}$ , except in Umeå, Sweden, where samples were stored in 1.8 mL plastic tubes in  $-80^{\circ}\text{C}$  freezers. Participants completed self-administered questionnaires on lifestyle factors and diet.

#### Follow-up for cancer incidence

Incident cancer cases were identified at regular intervals through population-based cancer registries (Denmark, Italy except Naples, the Netherlands, Norway, Spain, Sweden and United Kingdom) or by active follow-up (France, Germany, Greece and Naples), which involved a combination of methods, including review of health insurance records, cancer and pathology registries, as well as direct contact with participants and their next-of-kin.

Mortality data, including vital status, cause of death, and date of death, were obtained from mortality registries at the regional or national level. Subjects were followed up from study entry until cancer diagnosis (except non-melanoma skin cancer), death, emigration, or the end of the follow-up period for the relevant study centre. End of follow-up was defined as the latest date of complete follow-up for both cancer incidence and vital status and varied between study centres from December 2004 to June 2010. Vital status at follow-up is over 98% complete.

#### Selection of case and control participants

We initially identified 1,273 subjects diagnosed with incident head and neck or esophagus cancer within the entire EPIC cohort by the end of the follow-up period for all centres. These cancer cases were defined on the basis of the *International Classification of Diseases for Oncology, Second Edition* (ICD-O-2), and included: oral cavity (ICD C02.0-C02.9, C04.0-C04.9, C03.0-C03.9, C05.0-C06.9, C14.0-C14.9), oropharynx (C01.9, C02.4, C09.0-C10.9), hypopharynx (C13.0-C13.9), larynx (C32.0-C32.9), and esophagus (C15.0-C15.9). Cases who did not donate a blood sample ( $n = 152$ ), did not have enough plasma available for biochemical analysis ( $n = 20$ ), had a history of another cancer ( $n = 158$ , except non melanoma skin cancer), were not histologically confirmed, were prevalent at the time of blood donation, or did not

have questionnaire information available ( $n = 22$ ), were excluded, leaving 921 eligible cases. Because the etiology of squamous cell carcinoma (SCC) and adenocarcinoma are thought differ and the vast majority of HNC are SCC, we excluded adenocarcinoma of the head and neck ( $n = 16$ ). However, because esophagus cancers are more evenly split between adenocarcinoma and SCC, we retained all esophagus cancer cases, but focus primarily on the SCC in our analysis. After excluding cases from Denmark ( $n = 288$ ) and the Malmö centre in Sweden ( $n = 101$ ) who did not participate in this study, 516 eligible case participants with plasma samples were available for biochemical analyses. Data on histology were collected from each centre where possible.

For each case participant, one control was randomly chosen from appropriate risk sets consisting of all cohort members alive and free of cancer (except non-melanoma skin cancer) at the time (and hence age) of diagnosis of the index case. Matching criteria were country, sex, date of blood collection ( $\pm 1$  month, relaxed to  $\pm 5$  months for sets without available controls), and date of birth ( $\pm 1$  year, relaxed to  $\pm 5$  years for sets without available control participants). In addition, we included 479 additional controls (control group 2) that were analysed in the context of a parallel study and individually matched to cases of another cancer site using identical matching criteria.

The final dataset included 516 cancer cases and 516 individually matched controls, as well as 479 additional unmatched controls from control group 2 that contributed to unconditional and stratified risk analyses.

#### Biochemical analyses

All biochemical analyses were performed at Bevital A/S (<http://www.bevital.no>), Bergen, Norway. The study included measurements of plasma levels of B2 (riboflavin), B6 (measured as pyridoxal 5'-phosphate, its active form), folate (B9), B12 (cobalamin), total homocysteine, and methionine. All case and control participants were successfully analyzed, for at least one of the biomarker. We also measured cotinine as an indicator of recent smoking behavior. Levels of B2, B6, homocysteine, methionine, and cotinine were determined by mass spectrometry-based methods (liquid chromatography coupled to tandem mass spectrometry; gas chromatography coupled to tandem mass spectrometry)<sup>17,18</sup> and microbiological methods were used to determine levels of folate (*Lactobacillus casei*)<sup>19,20</sup> and B12 (*Lactobacillus leichmannii*).<sup>21</sup>

Samples were analysed in batches of 86 and quality control included six calibration samples, two control samples, and one blank sample in each batch. The coefficients of variations (CVs) within and between batches were, respectively, 6 and 11% for vitamin B2, 3% and 6% for vitamin B6, 4% and 5% for folate and vitamin B12, 1% and 3% for methionine, and 1% and 2% for homocysteine. All plasma samples were kept at  $-80^{\circ}\text{C}$  and all HNC and esophagus cancer cases and their individual matched controls were analyzed together within the same batches in random order, as were samples

from control group 2. The laboratory staff was blinded to the case-control status of the blood samples.

#### Statistical analyses

The relation of lifestyle and dietary factors with biomarker levels were assessed using linear regression models, adjusted for age, sex, and country.

Risk analysis involved calculating quartiles of plasma levels for each biomarker of interest based on their distribution among the matched control participants. Odds ratios (OR) and 95% confidence intervals (CI) for participants in the second, third, and fourth quartiles were calculated relative to the first quartile using conditional logistic regression, conditioning on individual case sets. Additional adjustment was conducted for indicators of risk factors, including smoking status [never, former, current, missing] and for quartiles of cotinine levels [defined by the distribution among current smokers], which was considered to be the most accurate measure of smoking intensity at the time of blood collection, as well as educational attainment (in four categories) and alcohol consumption at the time of recruitment (g/day). Adjusting for additional smoking variables (duration of smoking, average cigarettes smoked per day) did not alter the results notably and were not included in the final models.

In order to increase the statistical power and further evaluate the consistency of any association observed in the conditional logistic regression, we conducted a risk analysis including the 479 additional controls that were available using unconditional logistic regression adjusted for matching factors, including age at recruitment, sex, country, as well as risk factors using the same approach as in the conditional logistic regression outlined above.

Stratified analysis by tumor site was also conducted using unconditional logistic regression adjusting for matching factors (age at recruitment, sex, country), as well as risk factors (educational attainment, cotinine, smoking status and alcohol consumption at the time of recruitment).

As a measure of statistical significance for each biomarker and statistical model, we included the base 2 logarithm ( $\log_2$ ) of the biomarker levels as a continuous variable in a separate logistic regression model to estimate the  $p$  for trend. This model was also used in exploratory stratified risk analysis by pre-defined demographic characteristics and risk factors, and the OR trend estimate from this model ( $\log_2$ OR) may be interpreted as the relative risk associated with a doubling in plasma levels of the biomarker of interest. We used  $\chi^2$  tests to assess heterogeneity in  $\log_2$ OR estimates in stratified analyses.

All-cause mortality for differences in biomarker levels was evaluated among HNC and esophagus squamous cell carcinoma (ESCC) cases by Cox proportional hazard regression analyses using years since diagnosis as the time variable. Hazard ratios (HR) of all-cause mortality were calculated after adjusting for age at diagnosis, sex and country.

All  $p$  values were two-sided and statistical analyses were conducted using SAS version 9.2 (Cary, NC).

## Results

### Baseline characteristics

The final study population included 350 head and neck cancers and 166 esophagus cancers, and 516 individually matched controls. 68% of the nested case-control population were male and 32% were female (Table 1). The median age at recruitment was 62 years (5th–95th percentile: 49–77) and the average time from blood draw to diagnosis for cases was 6.4 years. Control group 2 included an additional 479 subjects with similar demographic characteristics as the matched control group, but with a higher proportion of women.

### Correlates of one-carbon metabolism biomarkers

The relation of dietary intake of major food groups, smoking, and alcohol intake with biomarker levels are presented in Supporting Information Table 1. Dietary intake of vegetable, of dairy products and meat products were inversely associated with homocysteine. Current smokers also had a notably lower levels of riboflavin ( $p < 10^{-6}$ ), vitamin B6 ( $p < 10^{-3}$ ) and folate ( $p < 10^{-3}$ ), than did never smokers. Additionally, correlation coefficients between the measured biomarkers are presented in Supporting Information Table 2.

### Plasma one-carbon metabolism biomarkers in relation to HNC and esophagus cancer risk

We conducted the main risk analysis for HNC and ESCC separately (Table 2). After adjusting for education, alcohol consumption, smoking status and cotinine, study participants with higher levels of homocysteine had an elevated risk of HNC, the OR when comparing the top and bottom quartiles [ $OR_{Q4 \text{ vs. } Q1}$ ] being 2.13 (95% confidence interval [95% CI], 1.13–4.00,  $p$  for trend 0.009). A decrease in risk was also seen among subjects with higher levels of folate ( $OR_{Q4 \text{ vs. } Q1}$  0.63, 95% CI 0.35–1.16,  $p$  for trend 0.02). Plasma levels of the different biomarkers displayed weak or no evidence of association with ESCC risk ( $p$  for trend  $> 0.06$ ). Including the additional controls in an unconditional logistic regression did not notably affect the association between homocysteine and HNC risk ( $p$  for trend 0.002), but the association of folate was attenuated and no longer statistically significant ( $p$  for trend 0.39). When comparing HNC cases with control group 2 only, homocysteine displayed similar associations as in the conditional analysis ( $OR_{Q4 \text{ vs. } Q1}$  2.30, 95% CI 1.47–3.59,  $P$  for trend  $3 \times 10^{-4}$ , and  $OR_{Q4 \text{ vs. } Q1}$  1.70, 95% CI 1.06–2.78,  $p$  for trend 0.03 for the minimally adjusted and fully adjusted model, respectively).

Stratified risk analysis by tumor site (Table 3) indicated that the association of homocysteine was particularly prominent for cancers of the oral cavity and gum ( $P$  for trend  $8 \times 10^{-4}$ ) and oropharynx cancer ( $p$  for trend 0.008), whilst no association was seen for ESCC, larynx and hypopharynx cancer or for adenocarcinoma of the esophagus ( $p$  for trend  $> 0.09$ ).

Overall, a doubling in plasma homocysteine was associated with a 53% higher odds of HNC and ESCC combined

Table 1. Baseline and clinical characteristics of study participants

	No. (%) participants in group		
	Cases (n = 516)	Matched controls (n = 516)	Control group n 2 (n = 479)
Discrete variables			
Participating countries			
France	7 (1)	7 (1)	13 (3)
Italy	65 (13)	65 (13)	93 (19)
Spain	93 (18)	93 (18)	59 (12)
United Kingdom	126 (24)	126 (24)	71 (15)
The Netherlands	71 (14)	71 (14)	52 (11)
Greece	21 (4)	21 (4)	18 (4)
Germany	98 (19)	98 (19)	131 (27)
Sweden	34 (7)	34 (7)	37 (8)
Norway	1 (0)	1 (0)	5 (1)
Sex			
Men	353 (68)	353 (68)	255 (53)
Women	163 (32)	163 (32)	224 (47)
Smoking status			
Never	105 (20)	214 (41)	220 (46)
Former	145 (28)	184 (36)	160 (33)
Years since quitting <10	55 (40)	47 (26)	44 (28)
Years since quitting ≥10	84 (60)	131 (74)	113 (72)
Current	256 (50)	104 (20)	96 (20)
Unknown	10 (2)	14 (3)	3 (1)
Education			
None	36 (7)	31 (6)	33 (7)
Primary school	191 (38)	170 (34)	158 (33)
Technical/professional school	114 (22)	136 (27)	104 (22)
Secondary school	78 (15)	64 (13)	64 (13)
Higher education	72 (14)	95 (19)	105 (22)
Unknown	18 (4)	9 (2)	13 (3)
Alcohol intake at recruitment (g/d)			
=0	89 (17)	60 (12)	57 (12)
0.1–6	136 (26)	167 (32)	175 (37)
6.1–12	48 (9)	77 (15)	58 (12)
12.1–24	71 (14)	92 (18)	90 (19)
24.1–60	109 (21)	94 (18)	81 (17)
60.1–96 in men or >60 in women	46 (9)	19 (4)	15 (3)
>96 in men	15 (3)	7 (1)	3 (1)
Body mass index <sup>1</sup>			
<18.5	9 (2)	5 (1)	1 (0)
18.5–25	192 (37)	201 (39)	190 (40)
25–30	230 (45)	239 (46)	207 (43)
≥30	85 (16)	71 (14)	81 (17)
Continuous variables, median (5th-95th percentile)			
Age at blood draw (yrs)	57 (42–71)	57 (42–71)	57 (41–68)

Table 1. Baseline and clinical characteristics of study participants (Continued)

	No. (%) participants in group		
	Cases (n = 516)	Matched controls (n = 516)	Control group n 2 (n = 479)
Serum levels for components of the 1-carbon metabolism			
Riboflavin, nmol/L	13.4 (5.3–49.9)	13.1 (5.9–47.9)	13.8 (5.4–44.4)
Pyridoxal phosphate, nmol/L	33.1 (13.9–101)	34.5 (14.2–98.8)	34.1 (14.2–89.2)
Serum folate, nmol/L	12.5 (3.5–41.3)	12.9 (5.1–36.4)	12.1 (4.3–38.5)
Cobalamin, Vit B12, pmol/L	319 (173–577)	328 (176–521)	337 (186–588)
Homocysteine, $\mu$ mol/L	10.8 (6.7–23.9)	10.2 (6.5–18)	9.8 (6.0–17.7)
Methionine, $\mu$ mol/L	25.1 (16.6–37.2)	25.1 (17.2–37.9)	24.6 (16.8–35.1)
Clinical characteristics, case participants only			
Age at diagnosis, median (range), yrs	62 (49–77)		
Time from blood draw to diagnosis	6.4 (0.75–12.8)		
Tumour site, no. (%)			
Esophagus	166 (32)		
Hypopharynx + larynx	145 (28)		
Gum + oral cavity	110 (21)		
Oropharynx	67 (13)		
Other	28 (6)		

<sup>1</sup>Body mass index is calculated as weight in kilograms divided by height in meters squared.

(OR for  $\log_2$  homocysteine [ $OR_{\log_2}$ ]; 1.53, 95% CI: 1.17–1.98,  $p$  for trend 0.001). The relation between homocysteine and squamous cell HNC/ESCC cancer by demographic characteristics and risk factors did not display any clear effect modifications in stratified analysis ( $p$  for heterogeneity >0.21) (Fig. 1 and Supporting Information Fig. 1). Notably, the inverse association of homocysteine with risk was observed both among never smokers ( $OR_{\log_2}$  1.88, 95% CI 1.08–3.28), as well as among current smokers ( $OR_{\log_2}$  1.56, 95% CI 1.03–2.35). Stratified risk analyses were also conducted for vitamin B2 (Supporting Information Fig. 2), B6 (Supporting Information Fig. 3), folate (Supporting Information Fig. 4), B12 (Supporting Information Fig. 5) and methionine (Supporting Information Fig. 6).

#### All-cause mortality for study participants diagnosed with HNC and esophagus cancer

Results of cox-proportional hazards regression models of all-cause mortality based on 277 deaths are shown in Supporting Information Table 3. Overall, plasma levels of the investigated OCM biomarkers did not display strong associations with all-cause mortality, nor with HNC-specific cause mortality (94 deaths). However, both cobalamin and methionine were weakly associated with all-cause mortality among HNCs in the unadjusted analysis, the hazard ratio (HR) when comparing the top and bottom quartiles [ $HR_{Q4 \text{ vs. } Q1}$ ] being 0.60 (95% CI: 0.38–0.96,  $p$  for trend 0.06) and 0.69 (95% CI: 0.44–1.11,  $p$  for trend 0.01), respectively, and adjusting for

potential confounders did not notably affect the HR estimates (Supporting Information Table 3).

#### Discussion

This is the first study investigating pre-diagnostic biomarkers of the one-carbon metabolism in relation to cancers of the head and neck and esophagus. We measured plasma levels of several B-vitamins and related metabolites, and observed that subjects with elevated levels of homocysteine and low levels of folate had higher risk of HNC. The other investigated biomarkers, including methionine and vitamins B2, B6, folate and B12, were not associated with risk of HNC or ESCC.

#### Homocysteine and cancers of the head and neck and esophagus

Only a few retrospective case-control studies have reported on circulating OCM biomarkers and HNC risk, most indicating lower folate levels and higher homocysteine levels among cases than among controls.<sup>22–25</sup> However, given the potential for reverse causality to influence the results in retrospective case-control studies, it is difficult to interpret those studies in the context of cancer etiology. We are not aware of any well-powered prospective studies investigating circulating one-carbon metabolism biomarkers and HNC cancer. This is noteworthy considering that head and neck cancers have been widely linked to dietary intake of fruits and vegetables that are important sources for specific B-vitamins involved in

Table 2. Odds ratios and 95% confidence intervals of head and neck cancer (HNC) and esophagus squamous cell carcinoma (ESCC) for plasma levels of vitamins B2, B6, folate, B12, homocysteine and methionine

Quartile (Range)	Case <sup>2</sup>	Control <sup>2</sup>	Odds ratio (95% confidence interval)							
			Head and neck cancer <sup>1</sup>				Esophagus squamous cell carcinoma			
			Matched controls unadjusted <sup>3</sup> (n = 350/350) <sup>3</sup>	Matched controls adjusted <sup>4</sup> (n = 350/350) <sup>4</sup>	All controls combined adjusted <sup>5</sup> n = (350/995) <sup>5</sup>	Matched controls unadjusted <sup>6</sup> (n = 73/73) <sup>6</sup>	Matched controls adjusted <sup>4</sup> (n = 73/73) <sup>4</sup>	All controls combined adjusted <sup>7</sup> (n = 73/995) <sup>7</sup>		
<b>Riboflavin (Vitamin B2), nmol/L<sup>8</sup></b>										
1 (2.5–9.4)	152 (29.5%)	129 (25.0%)	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
2 (9.5–13.1)	96 (18.6%)	129 (25.0%)	0.61 (0.40–0.94)	0.75 (0.45–1.27)	0.76 (0.51–1.13)	0.47 (0.16–1.36)	0.17 (0.02–1.26)	0.50 (0.21–1.19)		
3 (13.2–21.3)	145 (28.1%)	129 (25.0%)	1.01 (0.66–1.52)	1.33 (0.80–2.20)	1.12 (0.77–1.64)	0.66 (0.25–1.79)	0.57 (0.09–3.56)	1.05 (0.50–2.20)		
4 (21.4–199)	123 (23.8%)	129 (25.0%)	0.67 (0.42–1.05)	1.19 (0.68–2.08)	1.18 (0.77–1.79)	0.66 (0.24–1.79)	0.94 (0.18–4.92)	1.21 (0.54–2.72)		
<i>p</i> for trend <sup>9</sup>			0.21	0.22	0.31	0.33	0.84	0.41		
<b>Pyridoxal 5'-phosphate (Vitamin B6), nmol/L<sup>8</sup></b>										
1 (7.2–25.6)	160 (31%)	129 (25.0%)	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
2 (25.7–34.5)	116 (22.5%)	129 (25.0%)	0.67 (0.44–1.03)	0.75 (0.45–1.26)	0.92 (0.63–1.34)	0.88 (0.34–2.27)	1.96 (0.51–7.56)	1.09 (0.49–2.42)		
3 (34.6–47.6)	112 (21.7%)	129 (25.0%)	0.66 (0.43–1.02)	0.96 (0.57–1.63)	1.09 (0.73–1.61)	1.05 (0.38–2.92)	1.41 (0.28–7.18)	1.34 (0.59–3.04)		
4 (47.7–272)	128 (24.8%)	129 (25.0%)	0.77 (0.50–1.17)	1.01 (0.59–1.71)	1.02 (0.69–1.51)	1.31 (0.52–3.27)	1.82 (0.45–7.42)	2.26 (1.06–4.84)		
<i>p</i> for trend <sup>9</sup>			0.41	0.48	0.56	0.69	0.82	0.06		
<b>Serum folate (vitamin B9), nmol/L<sup>8</sup></b>										
1 (0.3–9.1)	163 (31.6%)	129 (25.0%)	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
2 (9.2–12.8)	108 (20.9%)	129 (25.0%)	0.51 (0.33–0.81)	0.62 (0.36–1.08)	0.80 (0.54–1.17)	1.08 (0.39–2.98)	7.78 (1.24–49.05)	1.04 (0.48–2.22)		
3 (12.9–18.1)	119 (23.1%)	129 (25.0%)	0.61 (0.39–0.98)	0.77 (0.44–1.36)	1.13 (0.77–1.67)	0.85 (0.31–2.36)	4.08 (0.74–22.60)	1.11 (0.49–2.49)		
4 (18.2–109)	126 (24.4%)	129 (25.0%)	0.48 (0.30–0.80)	0.62 (0.34–1.13)	0.92 (0.61–1.38)	1.18 (0.38–3.68)	12.45 (1.14–135.5)	1.03 (0.47–2.24)		
<i>p</i> for trend <sup>9</sup>			2 x 10 <sup>-4</sup>	0.02	0.39	0.44	0.14	0.69		
<b>Cobalamin (Vitamin B12), pmol/L<sup>8</sup></b>										
1 (75.1–265)	148 (28.7%)	129 (25.0%)	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
2 (266–328)	131 (25.4%)	129 (25.0%)	0.83 (0.54–1.27)	1.01 (0.61–1.68)	0.88 (0.61–1.29)	1.33 (0.43–4.10)	1.49 (0.33–6.73)	1.06 (0.50–2.24)		
3 (329–391)	92 (17.8%)	129 (25.0%)	0.58 (0.36–0.92)	0.59 (0.33–1.05)	0.66 (0.44–0.99)	0.75 (0.25–2.27)	1.07 (0.27–4.32)	0.81 (0.37–1.80)		
4 (392–2737)	145 (28.1%)	129 (25.0%)	0.91 (0.58–1.44)	1.20 (0.68–2.10)	0.90 (0.62–1.30)	1.10 (0.34–3.55)	0.94 (0.21–4.25)	1.07 (0.51–2.23)		
<i>p</i> for trend <sup>9</sup>			0.88	0.64	0.63	0.85	0.78	0.98		
<b>Homocysteine, μmol/L<sup>8</sup></b>										
1 (4.9–8.3)	109 (21.1%)	128 (24.8%)	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
2 (8.4–10.1)	123 (23.8%)	130 (25.2%)	1.35 (0.85–2.13)	1.28 (0.74–2.21)	1.26 (0.85–1.89)	0.80 (0.29–2.24)	0.52 (0.12–2.24)	1.07 (0.48–2.36)		

Epidemiology

Table 2. Odds ratios and 95% confidence intervals of head and neck cancer (HNC) and esophagus squamous cell carcinoma (ESCC) for plasma levels of vitamins B2, B6, folate, B12, homocysteine and methionine (Continued)

Quartile (Range)	Case <sup>2</sup>	Control <sup>2</sup>	Head and neck cancer <sup>1</sup>				Esophagus squamous cell carcinoma			
			Matched controls		All controls		Matched controls		All controls	
			unadjusted <sup>3</sup> (n = 350/350) <sup>4</sup>	adjusted <sup>4</sup> (n = 350/350) <sup>4</sup>	combined adjusted <sup>5</sup> n = (350/995) <sup>5</sup>	OR (95% CI)	unadjusted <sup>6</sup> (n = 73/73) <sup>6</sup>	adjusted <sup>6</sup> (n = 73/73) <sup>6</sup>	combined adjusted <sup>7</sup> (n = 73/995) <sup>7</sup>	OR (95% CI)
3 (10.2–12.4)	96 (18.6%)	129 (25%)	1.12 (0.70–1.79)	1.01 (0.58–1.76)	0.88 (0.58–1.36)	1.05 (0.37–3.00)	0.67 (0.16–2.71)	0.74 (0.31–1.74)		
4 (12.5–64.9)	188 (36.4%)	129 (25%)	2.81 (1.66–4.75)	2.08 (1.11–3.90)	1.81 (1.19–2.74)	1.67 (0.63–4.43)	1.09 (0.27–4.32)	1.69 (0.76–3.74)		
<i>p</i> for trend <sup>8</sup>			5 x 10 <sup>-5</sup>	0.009	0.002	0.07	0.52	0.11		
Methionine, µmol/L <sup>8</sup>										
1 (12.6–21.8)	149 (28.9%)	129 (25.0%)	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]		
2 (21.9–25.1)	110 (21.3%)	129 (25.0%)	0.72 (0.48–1.07)	0.91 (0.57–1.45)	0.83 (0.57–1.22)	0.78 (0.31–1.94)	1.59 (0.25–10.01)	0.76 (0.36–1.61)		
3 (25.2–29.3)	131 (25.4%)	129 (25.0%)	0.86 (0.56–1.32)	1.21 (0.71–2.06)	0.94 (0.64–1.38)	0.60 (0.22–1.64)	0.54 (0.10–3.00)	0.95 (0.44–2.04)		
4 (29.4–62.9)	126 (24.4%)	129 (25.0%)	0.85 (0.54–1.34)	1.30 (0.73–2.31)	0.99 (0.67–1.48)	0.74 (0.25–2.19)	1.23 (0.20–7.69)	1.29 (0.62–2.68)		
<i>p</i> for trend <sup>8</sup>			0.25	0.54	0.67	0.24	0.68	0.73		

<sup>1</sup>Adenocarcinoma excluded. <sup>2</sup>Numbers include all cancer cases of the head and neck and esophagus cancer cases, and individually matched control for whom laboratory measurements were available for complete case sets. Uninformative case-sets were excluded. <sup>3</sup>Assessed by analysing HNC cases and their individually matched controls by conditional logistic regression, conditioning on individual case set. <sup>4</sup>Further adjusted for educational attainment (in four groups), smoking status (never/former/current/missing), cotinine (quartiles defined among current smokers) and alcohol intake at recruitment (g/d). <sup>5</sup>Assessed by analysing HNC cases and all controls combined by unconditional logistic regression, adjusting for country, sex, age at recruitment (in 5-year groups), educational attainment (in four groups), smoking status (never/former/current/missing), cotinine (quartiles defined among current smokers) and alcohol intake at recruitment (g/day). <sup>6</sup>Assessed by analysing ESCC cases and their individually matched controls by conditional logistic regression, conditioning on individual case set. <sup>7</sup>Assessed by analysing ESCC cases and all controls combined by unconditional logistic regression, adjusting for country, sex, age at recruitment (in 5-year groups), educational attainment (in four groups), smoking status (never/former/current/missing), cotinine (quartiles defined among current smokers) and alcohol intake at recruitment (g/d). <sup>8</sup>Quartile cut-off points were determined based on the plasma level distribution of each biomarker among the 516 individually matched controls of the head and neck and esophagus. <sup>9</sup>*p* for trend assessed by the base 2 logarithm of the serum levels.



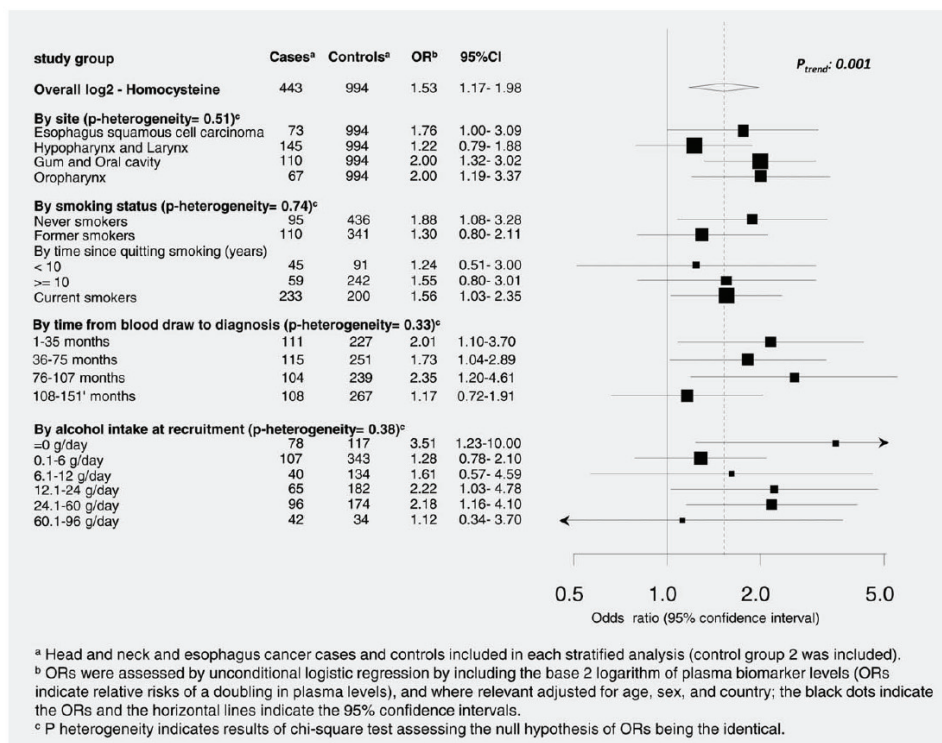


Figure 1. Forest plot showing stratified OR of cancer of the head and neck and esophagus for log2-of plasma homocysteine.

the pathways of one-carbon metabolism.<sup>2-4</sup> In the current EPIC study we observed a higher risk for HNC among subjects with elevated plasma levels of homocysteine. Most prospective studies linking specific cancer sites to OCM have implicated other factors with risk, in particular vitamin B6,<sup>12,15,26</sup> rather than homocysteine. Indeed, a clear decrease in risk of subsequent lung cancer with elevated levels of vitamin B6 and methionine was observed in a recent lung cancer study within EPIC,<sup>12</sup> whereas no associations of those biomarkers were observed with risk in the current study on HNC and ESCC. One observation that could explain this contrasting result is that the findings for B6 and methionine with lung cancer were particularly relevant for small cell carcinoma and adenocarcinoma, while no clear association was found for squamous cell carcinoma, the histology representing the majority of cancer cases of the head and neck and esophagus.

In the EPIC study population, the increase in HNC risk was primarily observed among subjects in the top quartile of plasma homocysteine who experienced approximately double the risk of those in quartiles 1 to 3 (Table 2). Apart

from larynx/hypopharynx cancer, this association was evident for all squamous cell cancers, including cancers of the esophagus, oral cavity and oropharynx, but not for adenocarcinoma of the esophagus (Table 3). When adjusting for potential confounders, including tobacco exposure and alcohol intake, the OR estimates were moderately attenuated (Table 2), but in stratified analysis the association of homocysteine with HNC and esophagus cancer risk was present both among non-alcohol consumers and non-smokers (Fig. 1), suggesting that residual confounding by these risk factors does not explain the association with risk. Further, while the association of homocysteine risk was particularly evident for HNC cancers that were diagnosed within three years after blood draw (Fig. 1), it was clearly present during the whole follow-up period (Fig. 1, Supporting Information Table 4). This observation indicates that diagnosis related behavioral or pre-malignant changes that may affect circulating biomarkers close to diagnosis does not explain the overall association of homocysteine with risk, i.e. as would be expected by reverse causality. We also compared homocysteine levels of HNC cases with an alternative control

Epidemiology

Table 3. Odds ratios and 95% confidence intervals of upper aero-digestive tract (UADT) cancer for plasma levels of vitamins B2, B6, folate, B12, homocysteine and methionine by tumour sites<sup>1</sup>

Quartile (range)	Control (n = 995)	Odds ratio (95% confidence interval)						
		Esophagus adenocarcinoma cases (n = 74)	Larynx and hypopharynx cases (n = 144) <sup>2</sup>	Oral cavity and Gum cases (n = 108) <sup>2</sup>	Oropharynx cases (n = 67) <sup>2</sup>			
<b>Riboflavin (vitamin B2), nmol/L<sup>3</sup></b>								
1 (2.5–9.4)	244 (24.6%)	1 [Reference]	55 (37.9%)	1 [Reference]	29 (26.9%)	1 [Reference]	21 (31.3%)	1 [Reference]
2 (9.5–13.1)	237 (23.8%)	1.36 (0.58–3.20)	29 (20%)	0.61 (0.35–1.08)	23 (21.3%)	1.01 (0.54–1.90)	11 (16.4%)	0.52 (0.22–1.22)
3 (13.2–21.3)	272 (27.3%)	2.2 (29.7%)	42 (29%)	0.92 (0.54–1.56)	31 (28.7%)	1.47 (0.80–2.72)	19 (28.4%)	1.04 (0.50–2.17)
4 (21.4–199)	242 (24.3%)	2.6 (35.1%)	19 (13.1%)	0.75 (0.39–1.44)	25 (23.1%)	1.63 (0.83–3.22)	16 (23.9%)	1.24 (0.55–2.78)
<i>p</i> for trend <sup>4</sup>		0.21		0.85		0.16		0.38
<b>Pyridoxal 5'-phosphate (Vitamin B6), nmol/L<sup>3</sup></b>								
1 (7.2–25.6)	259 (26%)	23 (31.1%)	1 [Reference]	44 (30.3%)	1 [Reference]	41 (38%)	1 [Reference]	22 (32.8%)
2 (25.7–34.5)	250 (25.2%)	18 (24.3%)	0.81 (0.40–1.61)	34 (23.4%)	1.00 (0.56–1.78)	29 (26.9%)	0.86 (0.49–1.50)	11 (16.4%)
3 (34.6–47.6)	232 (23.3%)	17 (23%)	0.88 (0.42–1.82)	28 (19.3%)	1.16 (0.62–2.14)	23 (21.3%)	0.80 (0.44–1.47)	19 (28.4%)
4 (47.7–272)	254 (25.5%)	16 (21.6%)	0.63 (0.30–1.33)	39 (26.9%)	1.34 (0.75–2.41)	15 (13.9%)	0.47 (0.24–0.91)	15 (22.4%)
<i>p</i> for trend <sup>4</sup>		0.58		0.65		0.17		0.05
<b>Serum folate (vitamin B9), nmol/L<sup>3</sup></b>								
1 (0.3–9.1)	269 (27.1%)	22 (29.7%)	1 [Reference]	43 (29.7%)	1 [Reference]	43 (39.8%)	1 [Reference]	20 (29.9%)
2 (9.2–12.8)	254 (25.6%)	13 (17.6%)	0.8 (0.36–1.80)	32 (22.1%)	1.02 (0.57–1.82)	22 (20.4%)	0.65 (0.36–1.18)	14 (20.9%)
3 (12.9–18.1)	223 (22.4%)	13 (17.6%)	0.86 (0.37–2.00)	43 (29.7%)	2.01 (1.12–3.60)	23 (21.3%)	0.84 (0.46–1.54)	14 (20.9%)
4 (18.2–109)	248 (24.9%)	26 (35.1%)	1.68 (0.79–3.56)	27 (18.6%)	1.05 (0.56–1.98)	20 (18.5%)	0.62 (0.33–1.18)	19 (28.3%)
<i>p</i> for trend <sup>4</sup>		0.24		0.99		0.04		0.77
<b>Cobalamin (Vitamin B12), pmol/L<sup>3</sup></b>								
1 (75.1–265)	239 (24%)	20 (27%)	1 [Reference]	43 (29.7%)	1 [Reference]	37 (34.3%)	1 [Reference]	20 (29.9%)
2 (266–328)	247 (24.8%)	23 (31.1%)	1.27 (0.64–2.52)	34 (23.4%)	0.89 (0.50–1.56)	27 (25%)	0.67 (0.38–1.20)	16 (23.9%)
3 (329–391)	228 (22.9%)	13 (17.6%)	0.79 (0.36–1.73)	28 (19.3%)	0.69 (0.38–1.23)	18 (16.7%)	0.50 (0.26–0.95)	14 (20.9%)
4 (392–2737)	280 (28.2%)	18 (24.3%)	1.17 (0.56–2.44)	40 (27.6%)	1.00 (0.57–1.74)	26 (24.1%)	0.64 (0.35–1.15)	17 (25.3%)
<i>p</i> for trend <sup>4</sup>		0.40		0.78		0.53		0.80
<b>Homocysteine, μmol/L<sup>3</sup></b>								
1 (4.9–8.3)	269 (27.1%)	10 (13.5%)	1 [Reference]	32 (22.1%)	1 [Reference]	19 (17.6%)	1 [Reference]	11 (16.4%)
2 (8.4–10.1)	238 (23.9%)	16 (21.6%)	1.35 (0.57–3.22)	34 (23.4%)	0.97 (0.53–1.79)	28 (25.9%)	1.64 (0.86–3.14)	16 (23.9%)
3 (10.2–12.4)	251 (25.3%)	19 (25.7%)	0.92 (0.38–2.18)	32 (22.1%)	0.67 (0.36–1.23)	18 (16.7%)	0.96 (0.46–2.00)	15 (22.4%)
4 (12.5–64.9)	236 (23.7%)	29 (39.2%)	1.19 (0.51–2.77)	47 (32.4%)	1.23 (0.67–2.26)	43 (39.8%)	2.45 (1.26–4.79)	25 (37.3%)
<i>p</i> for trend <sup>4</sup>		0.72		0.33		8 × 10 <sup>-4</sup>		0.008

Table 3. Odds ratios and 95% confidence intervals of upper aero-digestive tract (UADT) cancer for plasma levels of vitamins B2, B6, folate, B12, homocysteine and methionine by tumour sites<sup>1</sup>. (Continued)

Quartile (range) Methionine, $\mu\text{mol/L}$ <sup>2</sup>	Control (n = 995)	Odds ratio (95% confidence interval)				
		Esophagus adenocarcinoma cases (n = 74)	Larynx and hypopharynx cases (n = 144) <sup>3</sup>	Oral cavity and Gum cases (n = 108) <sup>2</sup>	Oropharynx cases (n = 67) <sup>2</sup>	
1 (12.6–21.8)	265 (26.6%)	21 (28.4%)	38 (26.2%)	31 (28.7%)	28 (41.8%)	1 [Reference]
2 (21.9–25.1)	255 (25.6%)	13 (17.6%)	32 (22.1%)	24 (22.2%)	16 (23.9%)	0.67 (0.34–1.32)
3 (25.2–29.3)	252 (25.4%)	23 (31.1%)	37 (25.5%)	31 (28.7%)	11 (16.4%)	0.50 (0.23–1.10)
4 (29.4–62.9)	223 (22.4%)	17 (23%)	38 (26.2%)	22 (20.4%)	12 (17.9%)	0.56 (0.26–1.22)
p for trend <sup>4</sup>		0.21	0.53	0.57		0.01

Assessed by analysing cancer cases of the head and neck and esophagus, and all controls combined by unconditional logistic regression, adjusting for country, sex, age at recruitment (in 5-year groups), educational attainment (in four groups), smoking status (never/former/current/missing), cotinine (quartiles defined among current smokers) and alcohol intake at recruitment (g/day). <sup>1</sup>Adenocarcinoma excluded. <sup>2</sup>Quartile cut-off points were determined based on the plasma level distribution of each biomarker among the 516 individually matched controls of the head and neck and esophagus. <sup>3</sup>p for trend assessed by the base 2 logarithm of the serum levels.

series in an unconditional analysis and the results on risk were consistent with that of the conditional analysis, thus further highlighting the consistency of the overall findings. In contrast, no convincing associations were observed between pre-diagnostically measured one-carbon metabolism biomarkers and all-cause mortality among cases, nor for cause-specific mortality considering events where HNC or ESCC was indicated as underlying cause of death (data not shown). Evaluating other specific causes of death, e.g. cardiovascular disease, was not feasible due to few relevant. Furthermore, it would have been informative to attain blood samples at diagnosis to fully evaluate the importance of these biomarkers for HNC survival.

**Potential roles of homocysteine in carcinogenesis**

Homocysteine is a thiol-containing amino acid produced through the catabolism of methionine.<sup>27</sup> Experiencing very high blood levels of homocysteine is a condition typically referred to as hyperhomocysteinemia, a well-established risk indicator for cardiovascular disease.<sup>28–30</sup> While randomized trials have shown that folate supplementation is an efficient means to reduce circulating homocysteine levels, it does not seem to translate into reduced risk of subsequent cardiovascular disease.<sup>31–33</sup> The importance of homocysteine in carcinogenesis is not clear. Homocysteine can be re-methylated to methionine by receiving a methyl group from folate, a reaction catalyzed by the vitamin B12-dependant methionine synthase. Some indication of carcinogenic effects of homocysteine have been provided by *in vitro* studies in which high homocysteine levels have been associated with increasing proliferation rates and oxidative stress.<sup>34,35</sup>

Previous studies investigating the association between circulating homocysteine and other cancer sites have not provided consistent results on risk. For instance, some prospective cohort studies have reported that elevated homocysteine was associated with increased risk of colorectal adenoma,<sup>36–39</sup> but not increased risk of subsequent colorectal cancer (CRC).<sup>40–43</sup> However, the largest and most recent prospective study on CRC indicated that high circulating homocysteine was associated with around 50% increased risk.<sup>44</sup> In our previous investigation of lung cancer,<sup>12</sup> we initially noted a positive association between homocysteine and risk in an unadjusted analysis, consistent with the well-established reverse relationship between folate and homocysteine. However, in contrast to the current study, tobacco smoking fully accounted for the association of homocysteine with risk in subsequent adjusted analysis (only the smoking adjusted analysis was reported).

Importantly, because folate deficiency leads to accumulation of homocysteine we cannot exclude the possibility that the positive association of homocysteine with risk observed in the current study indicates an underlying inverse relation between folate and risk (Supporting Information Table 2). In conditional analysis we observed a nominal inverse association between folate and risk that was primarily driven by an

increase in risk among subjects in the bottom quartile compared to those in quartiles 2 to 4. This observation would seem consistent with the inverse relation between folate and homocysteine, and that the increase in risk of homocysteine was primarily seen among subjects in the top quartile. Indeed, mutually adjusted conditional analysis (data not shown) indicated that these risk associations were not independent, homocysteine accounting for the association of folate, but not vice-versa. Nonetheless, the association between folate and HNC risk seen in conditional analysis was not supported by a comparison of the cases with the alternative control series. This result would not seem to provide strong support for a decrease in risk in subjects with high dietary intake of folate overall as seen in some previous studies,<sup>45,46</sup> nor in subgroups defined by head and neck cancer risk factors (Supporting Information Fig. 4).

Evaluating if the notably stronger and consistent risk association of homocysteine than of folate is reproducible will require an additional, well-powered prospective analysis on HNC and ESCC, a study that may not be feasible a single cohort given the rarity of the disease. Studies of relevant gene variants that are proxies for true exposures<sup>47</sup> may also provide information on the importance of maintaining adequate folate and moderate homocysteine levels over the life course along the lines of Mendelian randomization.<sup>48</sup>

## References

- Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst* 1981;66:1191–308.
- Chuang SC, Jenab M, Heck JE, et al. Diet and the risk of head and neck cancer: a pooled analysis in the INHANCE consortium. *Cancer Causes Control* 2012;23:69–88.
- Boeing H, Dietrich T, Hoffmann K, et al. Intake of fruits and vegetables and risk of cancer of the upper aero-digestive tract: the prospective EPIC-study. *Cancer Causes Control* 2006;17:957–69.
- Freedman ND, Park Y, Subar AF, et al. Fruit and vegetable intake and head and neck cancer risk in a large United States prospective cohort study. *Int J Cancer* 2008;122:2330–6.
- Olsen A, Halkjaer J, van Gils CH, et al. Dietary intake of the water-soluble vitamins B1, B2, B6, B12 and C in 10 countries in the European Prospective Investigation into Cancer and Nutrition. *Eur J Clin Nutr* 2009;63(Suppl 4):S122–49.
- United States Department of Health and Human Services USDA. Dietary Guidelines for Americans, U.S. Dept. of Health and Human Services, U.S. Dept. of Agriculture, Washington D.C. vol. 2005, 2005.
- Wcrf/Aicr, Food, nutrition, physical activity and the prevention of cancer: a global perspective, World Cancer Research Fund International, printed in United States of America. 2007.
- Choi SW, Mason JB. Folate and carcinogenesis: an integrated scheme. *J Nutr* 2000;130:129–32.
- Kim YI. Folate and carcinogenesis: evidence, mechanisms, and implications. *J Nutr Biochem* 1999;10:66–88.
- de Vogel S, Schneede J, Ueland PM, et al. Biomarkers related to one-carbon metabolism as potential risk factors for distal colorectal adenomas. *Cancer Epidemiol Biomarkers Prev* 2011;20:1726–35.
- Gibson TM, Weinstein SJ, Mayne ST, et al. A prospective study of one-carbon metabolism biomarkers and risk of renal cell carcinoma. *Cancer Causes Control* 2010;21:1061–9.
- Johansson M, Relton C, Ueland PM, et al. Serum B vitamin levels and risk of lung cancer. *JAMA* 2010;303:2377–85.
- Johansson M, Van Guelpen B, Vollset SE, et al. One-carbon metabolism and prostate cancer risk: prospective investigation of seven circulating B vitamins and metabolites. *Cancer Epidemiol Biomarkers Prev* 2009;18:1538–43.
- Schernhammer E, Wolpin B, Rifai N, et al. Plasma folate, vitamin B6, vitamin B12, and homocysteine and pancreatic cancer risk in four large cohorts. *Cancer Res* 2007;67:5553–60.
- Larsson SC, Orsini N, Wolk A. Vitamin B6 and risk of colorectal cancer: a meta-analysis of prospective studies. *Jama* 2010;303:1077–83.
- Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002;5:1113–24.
- Midttun O, Hustad S, Ueland PM. Quantitative profiling of biomarkers related to B-vitamin status, tryptophan metabolism and inflammation in human plasma by liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom* 2009;23:1371–9.
- Ueland PM, Midttun O, Windeberg A, et al. Quantitative profiling of folate and one-carbon metabolism in large-scale epidemiological studies by mass spectrometry. *Clin Chem Lab Med* 2007;45:1737–45.
- Molloy AM, Scott JM. Microbiological assay for serum, plasma, and red cell folate using cryopreserved, microtiter plate method. *Methods Enzymol* 1997;281:43–53.
- Ueland P. Folate in serum. [http://folk.uib.no/mfapu/Pages/BV/BVSite/Pages/analyte\\_info/s\\_folate.html](http://folk.uib.no/mfapu/Pages/BV/BVSite/Pages/analyte_info/s_folate.html). Accessed December 1 2009.
- Kelleher BP, Broin SD. Microbiological assay for vitamin B12 performed in 96-well microtitre plates. *J Clin Pathol* 1991;44:592–5.
- Almadori G, Bussu F, Galli J, et al. Serum folate and homocysteine levels in head and neck squamous cell carcinoma. *Cancer* 2002;94:1006–11.
- Almadori G, Bussu F, Galli J, et al. Serum levels of folate, homocysteine, and vitamin B12 in head and neck squamous cell carcinoma and in laryngeal leukoplakia. *Cancer* 2005;103:284–92.
- Eleftheriadou A, Chalastras T, Ferekidou E, et al. Association between squamous cell carcinoma of the head and neck and serum folate and homocysteine. *Anticancer Res* 2006;26:2345–8.
- Gorgulu O, Selcuk T, Ozdemir S, et al. Evaluation of the roles of serum vitamin B(12), folate and homocysteine levels in laryngeal squamous cell carcinoma. *J Int Med Res* 2010;38:2047–52.

## Conclusions

This study indicates that subjects with high levels of homocysteine are at increased risk of developing squamous cell carcinoma of the head and neck. Further large-scale prospective studies are warranted to confirm the robustness of homocysteine as an indicator of future risk, and its potential causal role in the pathogenesis of these cancers.

## Acknowledgements

Dr. Ueland reports that he is a member of the steering board of the nonprofit Foundation to Promote Research Into Functional Vitamin B12 Deficiency. No other disclosures were reported.

## Author contributions

M.J. and P.B. initiated, acquired the main funding, and designed this investigation. P.M.U., S.E.V. and Ø.M. led the laboratory analysis. A.F. conducted the statistical analysis under supervision of M.J. A.F. and M.J. drafted the first version of the manuscript with important contributions from P.B. All authors were involved with collection of data, data interpretation, critical revisions of the article, and approval of the final version. M.J. had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. E.R. is the overall coordinator of the EPIC study which he designed and implemented in collaboration with the main investigators in the collaborating centers.

26. Lurie G, Wilkens LR, Shvetsov YB, et al. Prediagnostic plasma pyridoxal 5'-phosphate (vitamin b6) levels and invasive breast carcinoma risk: the multiethnic cohort. *Cancer Epidemiol* 2012;21:1942-8.
27. Blom HJ, Smulders Y. Overview of homocysteine and folate metabolism. With special references to cardiovascular disease and neural tube defects. *J Inherit Metab Dis* 2011;34:75-81.
28. Homocysteine and risk of ischemic heart disease and stroke: a meta-analysis. *JAMA* 2002;288:2015-22.
29. Fassbender K, Mielke O, Hennerici M, Bertsch T. Plasma homocyst(e)ine concentrations in cerebrovascular disease. *Stroke* 1999;30:2244-5.
30. Wald DS, Law M, Morris JK. Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. *Br Med J* 2002;325:1202.
31. Cavalieri M, Schmidt R, Chen C, et al. B vitamins and magnetic resonance imaging-detected ischemic brain lesions in patients with recent transient ischemic attack or stroke: the VITamins TO Prevent Stroke (VITATOPS) MRI-substudy. *Stroke* 2012;43:3266-70.
32. Jardine MJ, Kang A, Zoungas S, et al. The effect of folic acid based homocysteine lowering on cardiovascular events in people with kidney disease: systematic review and meta-analysis. *Br Med J* 2012;344:e3533.
33. Zhou YH, Tang JY, Wu MJ, et al. Effect of folic acid supplementation on cardiovascular outcomes: a systematic review and meta-analysis. *PLoS One* 2011;6:e25142.
34. Sydow K, Schwedhelm E, Arakawa N, et al. ADMA and oxidative stress are responsible for endothelial dysfunction in hyperhomocyst(e)inemia: effects of L-arginine and B vitamins. *Cardiovasc Res* 2003;57:244-52.
35. Tyagi N, Sedoris KC, Steed M, et al. Mechanisms of homocysteine-induced oxidative stress. *Am J Physiol Heart Circ Physiol* 2005;289:H2649-56.
36. Bobe G, Murphy G, Rogers CJ, et al. Serum adiponectin, leptin, C-peptide, homocysteine, and colorectal adenoma recurrence in the Polyp Prevention Trial. *Cancer Epidemiol* 2010;19:1441-52.
37. Levine AJ, Grau MV, Mott LA, et al. Baseline plasma total homocysteine and adenoma recurrence: results from a double blind randomized clinical trial of aspirin and folate supplementation. *Cancer Epidemiol* 2010;19:2541-8.
38. Martinez ME, Giovannucci E, Jiang R, et al. Folate fortification, plasma folate, homocysteine and colorectal adenoma recurrence. *Int J Cancer* 2006;119:1440-6.
39. Martinez ME, Henning SM, Alberts DS. Folate and colorectal neoplasia: relation between plasma and dietary markers of folate and adenoma recurrence. *Am J Clin Nutr* 2004;79:691-7.
40. Le Marchand L, White KK, Nomura AM, et al. Plasma levels of B vitamins and colorectal cancer risk: the multiethnic cohort study. *Cancer Epidemiol* 2009;18:2195-201.
41. Lee JE, Li H, Giovannucci E, et al. Prospective study of plasma vitamin B6 and risk of colorectal cancer in men. *Cancer Epidemiol* 2009;18:1197-202.
42. Van Guelpen B, Hultdin J, Johansson I, et al. Low folate levels may protect against colorectal cancer. *Gut* 2006;55:1461-6.
43. Weinstein SJ, Albanes D, Selhub J, et al. One-carbon metabolism biomarkers and risk of colon and rectal cancers. *Cancer Epidemiol* 2008;17:3233-40.
44. Miller JW, Beresford SA, Neuhauser ML, et al. Homocysteine, cysteine, and risk of incident colorectal cancer in the Women's Health Initiative observational cohort. *Am J Clin Nutr* 2013;97:827-34.
45. Pelucchi C, Talamini R, Negri E, et al. Folate intake and risk of oral and pharyngeal cancer. *Ann Oncol* 2003;14:1677-81.
46. Tavani A, Malerba S, Pelucchi C, et al. Dietary folates and cancer risk in a network of case-control studies. *Ann Oncol* 2012;23:2737-42.
47. Tanaka T, Scheet P, Giusti B, et al. Genome-wide association study of vitamin B6, vitamin B12, folate, and homocysteine blood concentrations. *Am J Hum Genet* 2009;84:477-82.
48. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32:1-22.

# **CHAPTER IV:**

## ***ONE-CARBON METABOLISM AND RENAL CELL CARCINOMA***

## Objective

In the previous chapter, we found that participants having higher circulating concentrations of homocysteine had elevated risk of HNC. We wanted then to evaluate the extent to which homocysteine or other biomarkers included in the OCM pathway may be related to the incidence and survival of RCC. An additional reason for studying the association between OCM biomarkers and RCC was that this specific cancer site had less strength of association with smoking compared to HNC.

## Selection of cases and controls

As for the HNC study, the RCC case-control study nested within the EPIC cohort used similar selection methods and further study design information has been described previously in Chapter II (see *Material and Methods, Study population and data collection*).

In this study, we initially identified 905 cases from 10 participating countries within EPIC that were diagnosed with RCC as C64.9 according to the International Classification of Diseases for Oncology, Second Edition (ICD-O-2). We excluded prevalent cases and cases with a prior history of another cancer ( $n = 85$ , except non melanoma skin cancer), cases who did not donate a blood sample ( $n = 153$ ), were not histologically confirmed ( $n = 27$ ), did not have questionnaire information available ( $n = 6$ ), and cases from the Malmö centre that did not participate in this study ( $n = 64$ ), leaving 570 eligible RCC cases.

As in HNC study, one control was randomly chosen for each case in RCC study, from appropriate risk sets consisting of all cohort members alive and free of cancer (except non-melanoma skin cancer) at the time (and hence age) of diagnosis of the index case. The matching criteria is the same of the HNC studies being country, sex, date of blood collection ( $\pm 1$  month, relaxed to  $\pm 5$  months for sets without available controls), and date of birth ( $\pm 1$  year, relaxed to  $\pm 5$  years for sets without available control participants).

As previously explained for the HNC study, we also included a control group 2 ( $n=553$ ) unmatched to the RCC incident cases but matched to cases of the HNC study (Chapter III). Similar matching criteria for the control group 2 compared to the RCC matched controls were also used. This control group 2 participated to unconditional and stratified analyses to assess the consistency of any association and increase the statistical power.

## Replication study

The first analyses from the EPIC sub-sample in the RCC study showed strong inverse association between circulating vitamin B6 and risk of renal cell carcinoma. In order to determine if this association was restricted to the EPIC study population, a replication study was conducted in a separate cohort, the Melbourne Collaborative Cohort Study (MCCS). The MCCS is a prospective cohort study of 41,514 women and men aged 40 to 79 years at the time of recruitment which took place between 1990 and 1994 in Melbourne, Australia. Following blood draw, plasma fractions were stored in liquid nitrogen tanks at -196°C. Incident cancer diagnoses were identified via routine linkage to the Victorian Cancer Registry (VCR). There is a statutory requirement that all cancers diagnosed in the state of Victoria (the state in which Melbourne is situated), excluding non-melanoma skin cancer be reported to the VCR. Vital status and cause of death information were obtained via linkage to the National Death Index of Australia.

The nested case-control study from MCCS counted a total of 144 case-control matched pairs. We ensured that incident cases and controls were selected using the same protocol as the EPIC cohort. As a note, data from the MCCS study were used as an independent replication analysis and then were not pooled with the EPIC data.

## Biochemical analyses

Biochemical procedures used in this study are identical to the HNC study (Chapter II), and details of the biochemical analyses have been provided previously in Chapter II (see *Material and Methods, Biochemical analyses*).

## Statistical analyses

In this study, we used similar statistical methods to the HNC study including conditional and unconditional logistic regression to investigate the association of OCM biomarkers with RCC risk. Hazard ratios for all-cause mortality for RCC cases were calculated using Cox proportional hazard regression models. Model based estimates of the survival function by biomarker quartiles were calculated using flexible parametric survival models. These calculations were performed using SAS 9.2 and Stata 12.1 for Linux. More details for statistical analyses are provided in Chapter II (see *Material and Methods, Statistical methods*).



# Article 2

## Circulating biomarkers of one-carbon metabolism in relation to renal cell carcinoma incidence and survival

Mattias Johansson\*, **Anouar Fanidi\***, David C. Muller\*, Julie K. Bassett\*, Øivind Midttun, Stein Emil Vollset, Ruth C. Travis, Domenico Palli, Amalia Mattiello, Sabina Sieri, Antonia Trichopoulou, , Pagona Lagiou, Dimitrios Trichopoulos, Börje Ljungberg, Göran Hallmans, Elisabete Weiderpass, Guri Skeie, Carlos A. González, Miren Dorronsoro, Petra H Peeters, H.B(as). Bueno-de-Mesquita, Martine M. Ros, Marie-Christine Boutron Ruault, Guy Fagherazzi, Françoise Clavel, María-José Sánchez, Aurelio Barricarte Gurrea, Carmen Navarro, J Ramon Quiros, Kim Overvad, Anne Tjønneland, Krassimira Aleksandrova, Paolo Vineis, Marc J. Gunter, Graham Giles, Caroline Relton, Elio Riboli, Heiner Boeing, Per Magne Ueland, Gianluca Severi, Paul Brennan

\*Authors contributed equally to this work

*Journal of National Cancer Institute.* 2014 Nov 5; 106: 1-11

## Circulating Biomarkers of One-Carbon Metabolism in Relation to Renal Cell Carcinoma Incidence and Survival

Mattias Johansson\*†, Anouar Fanidi\*, David C. Muller\*, Julie K. Bassett\*, Øivind Midttun, Stein Emil Vollset, Ruth C. Travis, Domenico Palli, Amalia Mattiello, Sabina Sieri, Antonia Trichopoulou, Pagona Lagiou, Dimitrios Trichopoulos, Börje Ljungberg, Göran Hallmans, Elisabete Weiderpass, Guri Skeie, Carlos A. González, Miren Dorronsoro, Petra H. Peeters, H. Blas), Bueno-de-Mesquita, Martine M. Ros, Marie-Christine Boutron Ruault, Guy Fagherazzi, Françoise Clavel, María-José Sánchez, Aurelio Barricarte Gurrea, Carmen Navarro, J. Ramon Quiros, Kim Overvad, Anne Tjønneland, Krassimira Aleksandrova, Paolo Vineis, Marc J. Gunter, Rudolf Kaaks, Graham Giles, Caroline Relton, Elio Riboli, Heiner Boeing, Per Magne Ueland†, Gianluca Severi†, Paul Brennan†

\* Authors contributed equally to this work.

† Senior authors.

Manuscript received December 31, 2013; revised April 11, 2014; accepted September 3, 2014.

**Correspondence to:** Mattias Johansson, PhD, Genetic Epidemiology Group, International Agency for Research on Cancer (IARC/WHO), 150 cours Albert Thomas, 69008 Lyon, France (e-mail: [JohanssonM@iarc.fr](mailto:JohanssonM@iarc.fr)).

**Background** The etiology of renal cell carcinoma (RCC) is only partially understood, but a metabolic component appears likely. We investigated biomarkers of one-carbon metabolism and RCC onset and survival.

**Methods** The European Prospective Investigation into Cancer and Nutrition (EPIC) recruited 385 747 participants with blood samples between 1992 and 2000, and this analysis included 556 RCC case-control pairs. A subsequent replication study included 144 case-control pairs nested within the Melbourne Collaborative Cohort Study (MCCS). Plasma concentrations of vitamin B2, vitamin B6, folate, vitamin B12, methionine and homocysteine were measured in prediagnostic samples and evaluated with respect to RCC risk using conditional and unconditional logistic regression models, and to all-cause mortality in RCC cases using Cox regression models. All statistical tests were two-sided.

**Results** EPIC participants with higher plasma concentrations of vitamin B6 had lower risk of RCC, the odds ratio comparing the 4<sup>th</sup> and 1<sup>st</sup> quartiles ( $OR_{4vs1}$ ) being 0.40 95% confidence interval [CI] = 0.28 to 0.57,  $P_{trend} < .001$ . We found similar results after adjusting for potential confounders (adjusted  $P_{trend} < .001$ ). In survival analysis, the hazard ratio for all-cause mortality in RCC cases when comparing the 4<sup>th</sup> and 1<sup>st</sup> quartiles ( $HR_{4vs1}$ ) of vitamin B6 was 0.57 (95% CI = 0.37 to 0.87,  $P_{trend} < .001$ ).

Subsequent replication of these associations within the MCCS yielded very similar results for both RCC risk ( $OR_{4vs1} = 0.47$ , 95% CI = 0.23 to 0.99,  $P_{trend} = .07$ ) and all-cause mortality ( $HR_{4vs1} = 0.56$ , 95% CI = 0.27 to 1.17,  $P_{trend} = .02$ ). No association was evident for the other measured biomarkers.

**Conclusion** Study participants with higher circulating concentrations of vitamin B6 had lower risk of RCC and improved survival following diagnosis in two independent cohorts.

JNCI J Natl Cancer Inst (2014) 106(12): dju327

The etiology of kidney cancer is not well understood, and there are notable unexplained differences in incidence. The highest rates worldwide are observed in the Czech Republic (1), and in the United States; the age-standardized rates (ASR) of kidney cancer are approximately two-fold higher for African Americans and European Americans than for Asians (2). Renal cell carcinoma (RCC) is the predominant type of kidney cancer and accounts for approximately 80% of cases (2).

Established risk factors for RCC include tobacco smoking, obesity, and hypertension, as well as recently discovered gene variants (2–4). Additionally, it has been suggested that diabetes mellitus may increase RCC risk, whereas lifestyle factors such as high physical activity, alcohol, and intake of fruits and vegetables may reduce risk (2). The relation between fruit and vegetable intake and RCC is intriguing and consistently observed in both retrospective and prospective case-control studies (5). While residual confounding

by tobacco smoking is of concern in the interpretation of these studies, a causal association cannot be excluded (6–8).

Fruits and vegetables are sources of B vitamins and other components of the one-carbon metabolism pathway, which is important in maintaining DNA methylation and DNA repair mechanisms in the body (9,10). Circulating concentrations of B-vitamins have been investigated in relation to multiple cancers, but only one prospective study has been published for RCC (11).

We sought to investigate whether concentrations of circulating B-vitamins and amino acids in the one-carbon metabolism pathway are related to RCC incidence and outcome using a large European cohort. To ensure the validity of our findings, we also conducted a replication of promising results within a separate Australian cohort.

## Methods

### Study Cohort—The European Prospective Investigation into Cancer and Nutrition (EPIC)

The EPIC study recruitment procedures have been previously described in detail (12), and important cohort information and follow-up procedures are provided in the [Supplementary Methods](#) (available online).

### Selection of Cases and Controls

We initially identified 905 cases within EPIC that were diagnosed with RCC as C64.9 according to the International Classification of Diseases for Oncology, Second Edition (ICD-O-2). After excluding prevalent cases and cases with a prior history of another cancer ( $n = 85$ , except nonmelanoma skin cancer), cases who did not donate a blood sample ( $n = 153$ ), were not histologically confirmed ( $n = 27$ ), did not have questionnaire information available ( $n = 6$ ), and cases from the Malmö center that did not participate in this study ( $n = 64$ ), 570 RCC cases remained eligible.

For each case, one control was randomly chosen from risk sets consisting of all cohort members alive and free of cancer (except nonmelanoma skin cancer) at the time of diagnosis of the index case. Matching criteria were: country, sex, date of blood collection ( $\pm$  one month, relaxed to  $\pm$  five months for sets without available controls), and date of birth ( $\pm$  one year, relaxed to  $\pm$  five years). Additionally, we included 553 controls (control group 2) matched to cases of a parallel study of head and neck cancer using identical matching criteria. Biochemical analyses were undertaken in the same laboratory under the same conditions, and at the same time for all cases and controls.

After excluding sets with only one case or control, 556 case-control sets remained, as well as 553 additional unmatched controls from control group 2 that contributed to unconditional and stratified analyses.

### Replication Study—The Melbourne Collaborative Cohort Study (MCCS)

To replicate promising associations, we designed a case-control study nested within the MCCS. Extensive details on recruitment and follow-up have been published previously (also see [Supplementary Methods](#), available online) (13,14). Incident cases and controls were selected using the same protocol as the EPIC

study. A total of 144 case-control pairs were available. The MCCS data were used as an independent replication analysis and were not pooled with the EPIC data.

### Biochemical Analyses

Plasma samples were sent on dry ice to the Bevitall A/S laboratory (<http://www.bevital.no>) in Bergen, Norway, where vitamin B2 (riboflavin), vitamin B6 (measured as pyridoxal 5'-phosphate, its active form), folate (vitamin B9), vitamin B12 (cobalamin), total homocysteine, and methionine were measured. Cotinine was measured as an indicator of recent smoking behavior. Details of the biochemical analyses are provided in [Supplementary Methods](#) (available online) (15–19).

### Statistical Analyses

Quartiles of plasma concentrations for each biomarker were calculated based on the distribution among controls. Odds ratios (ORs) and 95% confidence intervals (CIs) of RCC were calculated relative to the first quartile using conditional logistic regression, conditioning on individual case set. Log-linear trends ( $P_{\text{trend}}$ ) were calculated by including the base 2 logarithm ( $\log_2$ ) of the biomarker concentration as a continuous variable in separate models.

To assess the consistency of any association, we compared RCC cases with the additional unmatched controls (control group 2) using unconditional logistic regression, adjusting for sex, country, and age at recruitment (five-year groups). We also compared the RCC cases with all EPIC controls in order to increase the statistical power.

To evaluate whether known risk factors of RCC could explain any association, we assessed if OR estimates were affected after including indicator variables in the logistic regression models for tobacco smoking (smoking status at baseline [never, former, current] and quartiles of cotinine concentrations [determined by the distribution for current smokers]), alcohol intake at recruitment (g/day), lifetime alcohol intake [ever/never], obesity (indicated by body mass index [BMI], five categories: <18.5, 18.5–25, 25–30, 30–35,  $\geq 35$  kg/m<sup>2</sup>), educational attainment (in four categories), waist-to-hip ratio (quartiles), and self-reported history of hypertension (yes/no). Missing values for these covariates were imputed under a multivariable normal model. One hundred imputations were drawn from the imputation model, which included all covariates and biomarkers, as well as the case-control status. For binary and categorical covariates, we rounded imputed values to the nearest integer category (20,21). As a sensitivity analysis, we also fitted models excluding observations with missing covariate data.

To assess potential effect modification, we performed stratified analyses by country of recruitment, smoking status, age at diagnosis, sex, education, time from blood draw to diagnosis, and waist-to-hip ratio. These analyses included the  $\log_2$  of each biomarker as a continuous covariate in unconditional logistic regression models.  $\chi^2$  tests were applied to assess heterogeneity between ORs.

Hazard ratios (HRs) for all-cause mortality for RCC cases were calculated using Cox proportional hazards models. Time since diagnosis was used as the timescale, and all models were adjusted for age at diagnosis, sex, and country. Further adjustment was undertaken for tobacco smoking, alcohol intake at recruitment,

education, hypertension, and waist-to-hip ratio. Tests for trend were based on models including  $\log_2$  of the biomarker concentrations. Visual inspection of smoothed, scaled Schoenfeld residuals revealed no notable departure from proportional hazards. Model based estimates of the survival function by biomarker quartiles were calculated using flexible parametric survival models (22). Restricted cubic splines with five knots (placed at evenly spaced centiles of the uncensored log survival times) were used to model the baseline hazard.

Among the matched controls, partial correlation coefficients (conditional on country, age, and sex) were calculated between log-transformed biomarker concentrations and dietary intake of fruits, vegetables, meats, and dairy products, as well as alcohol intake, circulating cotinine, BMI, and waist-to-hip ratio. All  $P$  values were two-sided. Statistical analyses were conducted using SAS 9.2 (Cary, NC) and Stata 12.1 for Linux (Stata Corporation).

## Results

### The EPIC Study Population

Baseline and demographic characteristics of the EPIC study population are displayed in Table 1. Of the individually matched cases and controls, 56% were men and 44% were women. Median age at recruitment was 57 years (5<sup>th</sup>-95<sup>th</sup> percentile: 42–67), and average time from blood draw to RCC diagnosis was 6.7 years. Control group 2 included an additional 553 subjects with similar demographic characteristics as the matched control group, albeit with a higher proportion of men (68%). Known risk factors of RCC displayed expected differences between cases and controls, including waist-to-hip ratio, BMI, smoking status, and hypertension. No substantial correlations were observed between one-carbon metabolism biomarkers and dietary intake of major food groups, circulating cotinine, alcohol intake, BMI, or waist-to-hip ratio (Supplementary Table 1, available online).

### Plasma Concentrations of One-Carbon Metabolism Biomarkers and Risk of RCC

Initial conditional risk analysis was conducted by comparing cases and matched controls for quartiles of plasma biomarkers of one-carbon metabolism (Table 2). Concentrations of vitamin B2, folate, B12, homocysteine and methionine displayed weak or no evidence of association with RCC risk ( $P_{\text{trend}} > .06$ ). In contrast, participants with higher concentrations of vitamin B6 had a lower risk of RCC in a dose response fashion ( $P_{\text{trend}} < .001$ ), the OR when comparing the 4<sup>th</sup> and 1<sup>st</sup> quartiles ( $\text{OR}_{4\text{vs}1}$ ) being 0.40 (95% CI = 0.28 to 0.57). Adjusting for potential confounders did not notably affect the OR estimates (adjusted  $\text{OR}_{4\text{vs}1}$  = 0.43, 95% CI = 0.29 to 0.64,  $P_{\text{trend}} < .001$ ) (Table 2). In a sensitivity analysis in which participants with missing covariate information were excluded, the corresponding adjusted  $\text{OR}_{4\text{vs}1}$  was 0.34 (95% CI = 0.21 to 0.57,  $P_{\text{trend}} < .001$ ) (Supplementary Table 2, available online).

After accounting for vitamin B6, the other plasma biomarkers (vitamin B2, folate, B12, homocysteine and methionine) did not display any association with RCC risk (data not shown).

When comparing RCC cases with control group 2, we observed similar associations (Supplementary Table 2, available online).

After combining all EPIC controls in an unconditional analysis the unadjusted  $\text{OR}_{4\text{vs}1}$  was 0.43 (95% CI = 0.32 to 0.60,  $P_{\text{trend}} < .001$ ) and 0.49 (95% CI = 0.35 to 0.68,  $P_{\text{trend}} < .001$ ) after adjusting for risk factors (Supplementary Table 2, available online).

### Replication Study

In order to determine if the inverse relation between vitamin B6 and risk of RCC was restricted to the EPIC study population, we analyzed plasma concentrations of vitamin B6 and cotinine in 144 additional case-control pairs nested within the MCCS. The distribution of vitamin B6 within the MCCS was similar to that within EPIC; hence, the same quartile cutoff points were applied. We obtained very similar OR estimates of RCC for quartiles of vitamin B6 within the MCCS, the  $\text{OR}_{4\text{vs}1}$  being 0.47 (95% CI = 0.23 to 0.99), after adjusting for available risk factors (Table 3).

### Associations With a Doubling of Plasma Vitamin B6 Levels

A doubling in plasma vitamin B6 was associated with a 20% lower odds of RCC in MCCS (OR for  $\log_2\text{B6}$  [ $\text{OR}_{\log_2}$ ] = 0.80, 95% CI = 0.64 to 1.02,  $P_{\text{trend}} = .07$ ) and a 22% lower odds in EPIC ( $\text{OR}_{\log_2} = 0.78$ , 95% CI = 0.63 to 0.82,  $P_{\text{trend}} < .001$ ). The corresponding unadjusted  $\text{OR}_{\log_2}$  estimates after stratifying by various descriptive variables within EPIC are displayed in Figure 1. The association between concentrations of vitamin B6 and RCC risk was marginally more prominent for men than women, for current smokers than former and never smokers, and for subjects that were not hypertensive. The  $\text{OR}_{\log_2}$  for vitamin B6 among current smokers, after adjusting for hypertension, waist-to-hip ratio, educational attainment, alcohol intake, BMI, and circulating cotinine was 0.52 (95% CI = 0.39 to 0.71). Additional adjustment for number of cigarettes smoked per day and duration of smoking did not affect the estimate ( $\text{OR}_{\log_2} = 0.53$ , 95% CI = 0.39 to 0.72). We further note that the association between vitamin B6 and risk was evident when evaluating blood samples taken up to ten years prior to diagnosis (Figure 1). Circulating concentrations of vitamin B6 by demographic variables, risk factors, and tumor stage are provided in Supplementary Table 3 (available online).

### All-Cause Mortality for RCC Cases

Results of Cox proportional hazards regression for all-cause mortality (205 deaths in total) are shown in Table 4 and Supplementary Table 4 (available online). For vitamin B6 the HR for all-cause mortality for RCC cases when comparing the 4<sup>th</sup> and 1<sup>st</sup> quartiles was 0.57 (95% CI = 0.37 to 0.87) (Table 4). The corresponding  $P_{\text{trend}}$  was less than .001, and the trend HR for  $\log_2\text{B6}$  ( $\text{HR}_{\log_2}$ ) was 0.74 (95% CI = 0.62 to 0.88). This result was nearly identical when excluding cases diagnosed within two years of blood draw ( $\text{HR}_{\log_2} = 0.73$ , 95% CI = 0.60 to 0.90,  $P_{\text{trend}} = .003$ ). For RCC cause-specific mortality (147 deaths), the  $\text{HR}_{4\text{vs}1}$  was 0.36 (95% CI = 0.20 to 0.65), the corresponding  $\text{HR}_{\log_2}$  was 0.64 (95% CI = 0.52 to 0.79,  $P_{\text{trend}} < .001$ ). For comparison, we observed 36 deaths among the matched controls and the corresponding  $\text{HR}_{\log_2}$  of all-cause mortality was 1.01 (95% CI = 0.71 to 1.45,  $P_{\text{trend}} = .95$ ). Adjusting the all-cause mortality

**Table 1.** Baseline and demographic characteristics of the study participants from EPIC

Continuous variables	RCC cases	Matched controls	Control group 2
	No. (%), n = 556	No. (%), n = 556	No. (%), n = 553
Sex*			
Men	310 (56%)	310 (56%)	374 (67.6%)
Women	246 (44%)	246 (44%)	179 (32.4%)
Participating countries*			
France	13 (2%)	13 (2%)	7 (1.3%)
Italy	88 (16%)	88 (16%)	70 (13%)
Spain	52 (9%)	52 (9%)	100 (18%)
United Kingdom	67 (12%)	67 (12%)	130 (24%)
The Netherlands	46 (8%)	46 (8%)	77 (14%)
Greece	17 (3%)	17 (3%)	22 (4%)
Germany	125 (22%)	125 (22%)	104 (19%)
Sweden	32 (6%)	32 (6%)	41 (7%)
Denmark	112 (20%)	112 (20%)	
Norway	4 (1%)	4 (1%)	2 (0.4%)
Smoking status			
Never smokers	225 (41%)	244 (44%)	230 (42%)
Former smokers	160 (29%)	180 (32%)	199 (36%)
Years since quitting ≥10	94 (58%)	122 (68%)	135 (68%)
Years since quitting <10	66 (42%)	58 (32%)	64 (33%)
Current smokers	166 (30%)	129 (23%)	110 (20%)
Unknown	5 (1%)	3 (1%)	14 (3%)
Educational attainment			
Primary school	229 (41%)	206 (37%)	222 (40%)
Technical/professional school	124 (22%)	136 (25%)	141 (26%)
Secondary school	77 (14%)	66 (12%)	70 (13%)
Higher education	110 (20%)	134 (24%)	99 (18%)
Unknown	16 (3%)	14 (3%)	21 (4%)
Body mass index, kg/m <sup>2</sup>			
<18.5	2 (0%)	2 (0%)	5 (1%)
18.5–25	179 (32%)	221 (40%)	215 (39%)
25–30	247 (45%)	241 (43%)	258 (47%)
30–35	99 (18%)	69 (12%)	61 (11%)
≥35	29 (5%)	23 (4%)	14 (3%)
Waist-to-hip ratio			
0.56–0.79	92 (17%)	113 (20%)	86 (16%)
0.80–0.89	129 (23%)	145 (26%)	146 (26%)
0.90–0.94	103 (19%)	114 (21%)	128 (23%)
0.95–1.30	196 (35%)	148 (27%)	150 (27%)
Unknown	36 (7%)	36 (7%)	43 (8%)
Hypertension			
No	276 (50%)	325 (59%)	318 (58%)
Yes	192 (35%)	140 (25%)	115 (21%)
Unknown	88 (16%)	91 (16%)	120 (22%)
Alcohol intake			
Never drinkers	37 (7%)	22 (4%)	28 (5%)
Ever drinkers	509 (92%)	530 (95%)	520 (94%)
Unknown	10 (2%)	4 (1%)	5 (1%)
Alcohol intake at recruitment*			
<5 g/day	248 (45%)	221 (40%)	235 (43%)
5–20 g/day	156 (28%)	173 (30%)	168 (30%)
≥20 g/day	152 (27%)	162 (30%)	150 (27%)
<b>Continuous variables</b>		<b>Median (5th–95th percentile)</b>	
Age at recruitment, y	56.9 (42–67)	56.9 (41.8–67.3)	56.6 (41.0–70.6)
Age at diagnosis, y	63.7 (49–75)	-	-
Time from blood draw to diagnosis, y	6.7 (1–12)	-	-
Plasma concentrations for components of the one-carbon metabolism			
Vitamin B2 (Riboflavin), nmol/L	13.9 (5.2–47.8)	14.4 (5.8–48.0)	13.2 (5.9–48.1)
Vitamin B6 (Pyridoxal 5'-phosphate), nmol/L	30.3 (13.0–89.0)	35.9 (14.6–122)	34.8 (14.3–97.0)
Folate (Vitamin B9), nmol/L	11.3 (3.9–32.2)	11.9 (4.5–36.6)	13.0 (5.1–35.6)

(Table continues)

Downloaded from <http://jnci.oxfordjournals.org/> at CIRC IARC on December 2, 2014

Table 1. (Continued).

Continuous variables	RCC cases	Matched controls	Control group 2
	No. (%), n = 556	No. (%), n = 556	No. (%), n = 553
Vitamin B12 (cobalamin), pmol/L	332 (192–582)	344 (194–588)	329 (178–523)
Methionine, µmol/L	24.1 (16.3–38.1)	25.0 (16.9–37.0)	25.1 (17.2–37.9)
Homocysteine, µmol/L	9.6 (6.2–18.4)	9.9 (6.1–17.3)	10.1 (6.4–18.4)

\* Matching criteria were country, sex, and age at recruitment. EPIC = European Prospective Investigation into Cancer and Nutrition.

Table 2. Odds ratios of RCC for plasma concentrations of vitamins B2, B6, folate, B12, and homocysteine and methionine in the EPIC study

Quartile (range)	Cases, No. (%)	Controls, No. (%)	Odds ratio (95% confidence interval)	
			Unadjusted conditional risk analysis* (n = 556/556)*	Conditional risk analysis adjusted for risk factors† (n = 556/556)†
<b>Vitamin B2 (Riboflavin), nmol/L‡</b>				
1 (2.57–9.83)	164 (30)	139 (25)	1.00 (referent)	1.00 (referent)
2 (9.84–14.3)	132 (24)	139 (25)	0.80 (0.57 to 1.12)	0.81 (0.56 to 1.16)
3 (14.4–22.2)	133 (24)	139 (25)	0.81 (0.58 to 1.13)	0.79 (0.56 to 1.13)
4 (22.2–416)	127 (23)	139 (25)	0.77 (0.55 to 1.09)	0.81 (0.56 to 1.17)
<i>P</i> <sub>trend</sub> §			.13	.26
<b>Vitamin B6 (Pyridoxal 5'-phosphate), nmol/L‡</b>				
1 (5.95–25.4)	210 (38)	138 (25)	1.00 (referent)	1.00 (referent)
2 (25.4–35.9)	136 (25)	139 (25)	0.66 (0.48 to 0.91)	0.69 (0.49 to 0.98)
3 (35.9–51.9)	120 (22)	139 (25)	0.55 (0.39 to 0.77)	0.62 (0.43 to 0.89)
4 (51.9–436)	90 (16)	139 (25)	0.40 (0.28 to 0.57)	0.43 (0.29 to 0.64)
<i>P</i> <sub>trend</sub> §			<.001	<.001
<b>Vitamin B9 (Folate), nmol/L‡</b>				
1 (0.20–8.40)	152 (27)	138 (25)	1.00 (referent)	1.00 (referent)
2 (8.41–11.8)	138 (25)	139 (25)	0.88 (0.62 to 1.24)	0.96 (0.66 to 1.40)
3 (11.9–17.3)	134 (24)	139 (25)	0.85 (0.59 to 1.22)	0.96 (0.65 to 1.43)
4 (17.3–109)	131 (24)	139 (25)	0.81 (0.55 to 1.20)	0.84 (0.55 to 1.27)
<i>P</i> <sub>trend</sub> §			.08	.11
<b>Vitamin B12 (Cobalamin), pmol/L‡</b>				
1 (75.2–281)	154 (28)	138 (25)	1.00 (referent)	1.00 (referent)
2 (281–343)	148 (27)	139 (25)	0.96 (0.69 to 1.33)	0.96 (0.67 to 1.37)
3 (344–419)	137 (25)	139 (25)	0.88 (0.63 to 1.24)	0.87 (0.61 to 1.26)
4 (419–5000)	116 (21)	139 (25)	0.74 (0.52 to 1.05)	0.67 (0.46 to 0.99)
<i>P</i> <sub>trend</sub> §			.12	.06
<b>Methionine, µmol/L‡</b>				
1 (2.1–21.48)	151 (27)	138 (25)	1.00 (referent)	1.00 (referent)
2 (21.5–25)	166 (30)	140 (25)	1.03 (0.74–1.44)	1.07 (0.75 to 1.52)
3 (25.01–28.94)	115 (21)	140 (25)	0.70 (0.49–1.01)	0.70 (0.47 to 1.03)
4 (28.95–71.4)	124 (22)	138 (25)	0.77 (0.53–1.11)	0.80 (0.53 to 1.19)
<i>P</i> <sub>trend</sub> §			.06	.10
<b>Homocysteine, µmol/L‡</b>				
1 (3.68–8.03)	141 (25)	138 (25)	1.00 (referent)	1.00 (referent)
2 (8.04–9.87)	157 (28)	140 (25)	1.11 (0.79 to 1.56)	1.00 (0.69 to 1.43)
3 (9.88–11.9)	114 (21)	138 (25)	0.81 (0.56 to 1.17)	0.73 (0.49 to 1.09)
4 (12.0–64.9)	144 (26)	139 (25)	1.02 (0.69 to 1.51)	0.85 (0.56 to 1.30)
<i>P</i> <sub>trend</sub> §			.55	.98

\* Assessed by analyzing renal cell carcinoma (RCC) cases and their individually matched controls by conditional logistic regression, conditioning on individual case set. EPIC = European Prospective Investigation into Cancer and Nutrition.

† Assessed by analyzing RCC cases and their individually matched controls by conditional logistic regression after multiple imputation of missing covariate data, conditioning on individual case set, and adjusting for waist-to-hip ratio (quartiles defined among matched controls), hypertension (yes/no), educational attainment (four categories), smoking status (never/former/current), plasma cotinine (quartiles defined by the distribution for current smokers), alcohol intake at recruitment (g/day) and alcohol intake (ever/never). Case-control numbers only include those case sets where both the case and matched control had complete plasma measurements.

‡ Quartile cutoff points were determined based on the plasma level distribution of each biomarker for 556 individually matched controls.

§ *P*<sub>trend</sub> assessed by the base 2 logarithm of plasma concentrations. All statistical tests were two-sided.

Downloaded from <http://jnci.oxfordjournals.org/> at CIRC IARC on December 2, 2014

**Table 3.** Odds ratios of RCC for plasma concentrations of vitamin B6 for participants in the MCCS study

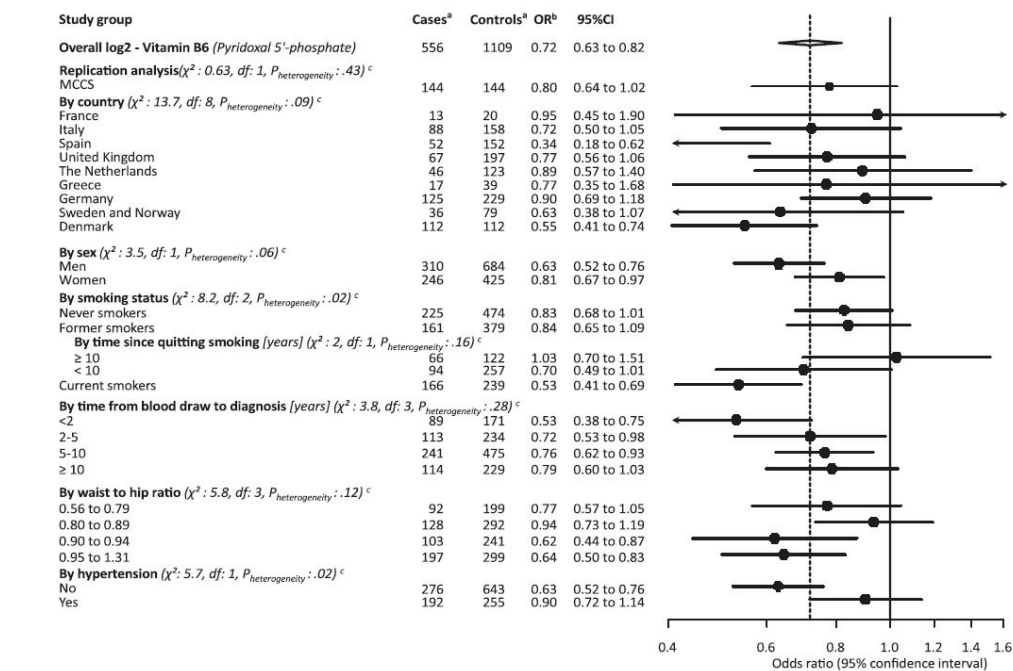
Quartile (range)	Cases, No. (%)	Controls, No. (%)	Odds ratio (95% confidence interval)	
			Unadjusted conditional risk analysis* (n = 144/144)*	Conditional risk analysis adjusted for risk factors† (n = 144/144)†
<b>Vitamin B6 (Pyridoxal 5'-phosphate), nmol/L ‡</b>				
1 (5.31–25.4)	47 (33)	35 (24)	1.00 (referent)	1.00 (referent)
2 (25.4–35.9)	40 (28)	31 (22)	0.86 (0.44 to 1.67)	0.80 (0.40 to 1.61)
3 (35.9–51.9)	28 (19)	40 (28)	0.46 (0.22 to 0.95)	0.48 (0.23 to 1.01)
4 (51.9–435)	29 (20)	38 (26)	0.50 (0.25 to 1.01)	0.47 (0.23 to 0.99)
<i>P</i> <sub>trend</sub> §			.02	.07

\* Assessed by analyzing renal cell carcinoma (RCC) cases and their individually matched controls by conditional logistic regression, conditioning on individual case set. MCCS = Melbourne Collaborative Cohort Study; RCC = renal cell carcinoma.

† Assessed by analyzing RCC cases and their individually matched controls by conditional logistic conditioning on individual case set, and adjusting for waist-to-hip ratio (continuous), smoking status (never/former/current), plasma cotinine (continuous), and alcohol intake at recruitment (g/day).

‡ Quartile cutoff points were determined based on the plasma level distribution of each biomarker for 556 individually matched controls in the European Prospective Investigation into Cancer and Nutrition study

§ *P*<sub>trend</sub> assessed by the base 2 logarithm of plasma concentrations. All statistical tests were two-sided.



**Figure 1.** Forest plot showing overall odds ratios of renal cell carcinoma (RCC) for the base 2 logarithm of plasma vitamin B6 for the European Prospective Investigation into Cancer and Nutrition (EPIC) study, the Melbourne Collaborative Cohort Study (MCCS), as well as further stratified within the EPIC study alone. a, RCC cases and controls included in each stratified analysis (control group 2 was included). b, Odds ratios (ORs) were assessed by unconditional logistic regression by including the base 2 logarithm of plasma concentrations (ORs

indicate relative risks of a doubling in plasma concentrations), and where relevant adjusted for age, sex, and country; the black dots indicate the ORs and the horizontal lines indicate the 95% confidence intervals. All statistical tests were two-sided. c, *P*<sub>heterogeneity</sub> indicates results of chi-square test assessing the null hypothesis of ORs being identical. The heterogeneity test for the replication analysis indicates any difference between the overall OR estimates of EPIC and MCCS. df = degrees of freedom.

analysis for potential confounders did not notably affect the HR estimates (Table 4). Model-based survival curves and Kaplan-Meier estimates are presented in Figure 2. Five-year survival

proportions were 57% (95% CI = 50% to 64%) for RCC cases in the bottom 25% of plasma vitamin B6 and 73% (95% CI = 62% to 81%) for RCC cases in the upper 25% of plasma vitamin

Downloaded from <http://jnci.oxfordjournals.org/> at CIRC IARC on December 2, 2014

**Table 4.** Hazard ratios of all-cause mortality for RCC cases for quartiles of plasma vitamins B2, B6, folate, B12, and homocysteine and methionine

Quartile (range)	Deceased*, No. (%)	Alive*, No. (%)	Person-years	Hazard ratio (95% confidence interval)	
				Minimally adjusted†	Adjusted for risk factors‡
<b>Vitamin B2 (Riboflavin), nmol/L§</b>					
1 (2.57–9.83)	50 (24)	111 (32)	744.1	1.00 (referent)	1.00 (referent)
2 (9.84–14.34)	52 (25)	75 (22)	602.8	1.23 (0.83 to 1.82)	1.32 (0.88 to 1.98)
3 (14.35–22.16)	48 (23)	85 (25)	528.9	1.01 (0.68 to 1.52)	1.09 (0.72 to 1.66)
4 (22.19–416.79)	55 (27)	73 (21)	501.4	1.19 (0.80 to 1.77)	1.38 (0.91 to 2.09)
<i>P</i> <sub>trend</sub> ¶				.75	.61
<b>Vitamin B6 (Pyridoxal 5'-phosphate), nmol/L§</b>					
1 (5.95–25.37)	90 (44)	116 (34)	890.4	1.00 (referent)	1.00 (referent)
2 (25.43–35.9)	51 (25)	83 (24)	576.5	0.87 (0.61 to 1.22)	0.87 (0.61 to 1.24)
3 (35.92–51.88)	37 (18)	84 (24)	530.8	0.64 (0.43 to 0.94)	0.66 (0.44 to 0.99)
4 (51.92–436.13)	27 (13)	61 (18)	379.5	0.57 (0.37 to 0.87)	0.59 (0.37 to 0.93)
<i>P</i> <sub>trend</sub> ¶				<.001	.004
<b>Vitamin B9 (Folate), nmol/L§</b>					
1 (0.2–8.4)	54 (26)	97 (28)	697.0	1.00 (referent)	1.00 (referent)
2 (8.41–11.84)	63 (31)	74 (22)	501.5	1.47 (1.02 to 2.13)	1.53 (1.04 to 2.24)
3 (11.86–17.25)	46 (22)	85 (25)	594.9	0.92 (0.61 to 1.37)	1.00 (0.66 to 1.51)
4 (17.27–109.35)	42 (20)	87 (25)	580.5	0.86 (0.57 to 1.30)	0.96 (0.62 to 1.47)
<i>P</i> <sub>trend</sub> ¶				.45	.88
<b>Vitamin B12 (Cobalamin), pmol/L§</b>					
1 (75.16–281.28)	61 (30)	90 (26)	638.9	1.00 (referent)	1.00 (referent)
2 (281.37–343.39)	57 (28)	90 (26)	600.5	0.94 (0.64 to 1.36)	0.96 (0.66 to 1.41)
3 (343.51–419)	46 (22)	88 (26)	595.5	0.75 (0.50 to 1.13)	0.75 (0.49 to 1.14)
4 (419.35–5000)	41 (20)	75 (22)	539.1	0.81 (0.54 to 1.21)	0.93 (0.61 to 1.41)
<i>P</i> <sub>trend</sub> ¶				.18	.58
<b>Methionine, µmol/L§</b>					
1 (2.1–21.48)	63 (31)	82 (24)	657.7	1.00 (referent)	1.00 (referent)
2 (21.5–25)	49 (24)	115 (33)	765.7	0.68 (0.46 to 0.99)	0.74 (0.50 to 1.10)
3 (25.01–28.94)	43 (21)	71 (21)	455.3	0.85 (0.57 to 1.27)	0.96 (0.63 to 1.44)
4 (28.95–71.4)	50 (24)	76 (22)	498.4	0.96 (0.65 to 1.41)	0.99 (0.67 to 1.47)
<i>P</i> <sub>trend</sub> ¶				.95	.67
<b>Homocysteine, µmol/L§</b>					
1 (3.68–8.03)	36 (18)	103 (30)	658.3	1.00 (referent)	1.00 (referent)
2 (8.04–9.87)	58 (28)	96 (28)	677.6	1.34 (0.88 to 2.05)	1.30 (0.85 to 2.00)
3 (9.88–11.94)	46 (22)	68 (20)	492.5	1.44 (0.92 to 2.25)	1.38 (0.88 to 2.16)
4 (11.95–64.88)	65 (32)	77 (22)	548.8	1.56 (1.02 to 2.40)	1.46 (0.94 to 2.25)
<i>P</i> <sub>trend</sub> ¶				.04	.11

\* Vital status for RCC case at the last follow-up.

† Assessed by analyzing renal cell carcinoma (RCC) cases by Cox-proportional hazards regression, adjusting for country, sex, and age at diagnosis. RCC = renal cell carcinoma.

‡ Assessed by analyzing RCC cases by Cox-proportional hazards regression after multiple imputation of missing covariate data, adjusting for country, sex, and age at diagnosis, and further by quartiles of vitamin B6 (cutoffs defined in controls), hypertension (yes/no), waist-to-hip ratio (quartiles), educational attainment (four categories), smoking status (never/former/current), plasma cotinine (quartiles defined by the distribution for current smokers), alcohol intake at recruitment (g/day) and body mass index (three categories were defined: <25 kg/m<sup>2</sup>; 25–30 kg/m<sup>2</sup>; >30 kg/m<sup>2</sup>, in order to have a reasonable distribution of body mass index in each group). A corresponding sensitivity analysis is also provided in Supplementary Table 3 (available online) where participants with missing covariate data were excluded.

§ Quartile cutoff points were determined based on the plasma level distribution of each biomarker for 556 individually matched controls.

¶ *P*<sub>trend</sub> assessed by the base 2 logarithm of plasma concentrations. All statistical tests were two-sided.

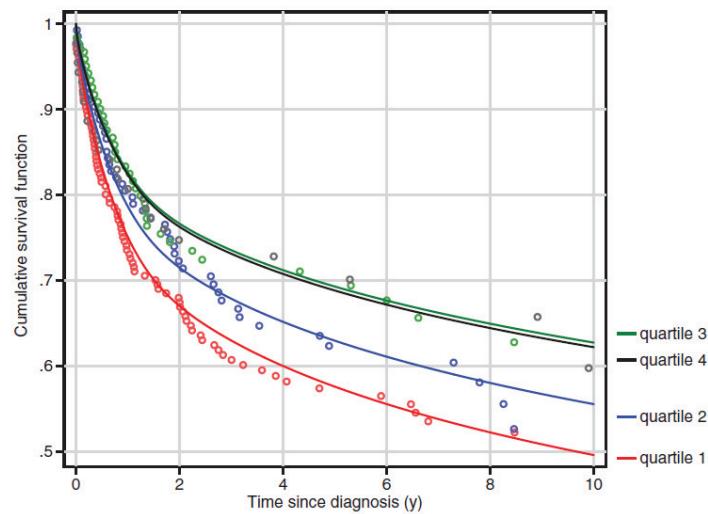
B6. Homocysteine also showed some association with all-cause mortality (*P*<sub>trend</sub> = .04), but adjusting for potential confounders attenuated this association.

We replicated the results of vitamin B6 with RCC survival in MCCS (57 deaths in total), the all-cause mortality HR<sub>4vs1</sub> estimate being 0.56 (95% CI = 0.27 to 1.17), and a doubling in vitamin B6 concentrations was associated with approximately 30% lower risk of death (HR<sub>log2</sub> = 0.69, 95% CI = 0.52 to 0.93, *P*<sub>trend</sub> = .02), and this estimate remained unchanged after adjusting for potential confounders

(HR<sub>log2</sub> = 0.69, 95% CI = 0.50 to 0.95, *P*<sub>trend</sub> = .02). Similarly as in EPIC, for RCC cause-specific mortality in the MCCS (36 deaths) the association was strengthened (HR<sub>log2</sub> = 0.54, 95% CI = 0.35 to 0.85, *P*<sub>trend</sub> = .008).

Within the EPIC case series, stage at diagnosis was only available for 45% of cases, though the characteristics of cases with and without information on tumor stage were similar (Supplementary Table 5, available online). Within the subset of cases with stage available, an association between B6 and





**Figure 2.** Cumulative survival curves of all-cause mortality in European Prospective Investigation into Cancer and Nutrition (EPIC) for study participants diagnosed with renal cell carcinoma (RCC) by quartiles of prediagnostic plasma vitamin B6. Cumulative survival curves of all-cause mortality for study participants diagnosed with RCC by prediagnostic

plasma concentrations of vitamin B6 (quartiles based on the distribution for the controls). The smooth lines depict survival functions calculated from a flexible parametric survival model with proportional hazards for vitamin B6 quartiles. The scattered points are Kaplan-Meier estimates of the survival functions, evaluated at the time of death for each failure.

mortality was evident before adjusting for stage ( $HR_{log_2} = 0.67$ , 95% CI = 0.50 to 0.91). As expected, adjusting for stage partly attenuated this association ( $HR_{log_2} = 0.81$ , 95% CI = 0.60 to 1.10). Similarly in MCCS for which stage was available for 94% of cases, adjusting for stage attenuated the HR estimates ( $HR_{log_2} = 0.81$ , 95% CI = 0.55 to 1.20).

## Discussion

We investigated if differences in prediagnostic circulating concentrations of B vitamins and additional biomarkers of the one-carbon metabolism were associated with differences in risk of RCC and related mortality. We found that subjects with higher plasma concentrations of vitamin B6 had a clear decrease in risk of subsequent RCC in two separate cohorts, as well as improved survival following diagnosis.

To date, only one prospective study assessing circulating concentrations of Vitamin B6 and RCC has been published (11). This study was conducted on 224 RCC case-control pairs nested within the Alpha-Tocopherol, Beta-Carotene (ATBC) Cancer Prevention Study and did not, in contrast to our finding, report an association between vitamin B6 and risk. In our current EPIC study based on 556 prospectively collected case-control pairs, higher vitamin B6 concentrations were clearly associated with reduced RCC incidence, with subjects in the top quartile having approximately half the risk of developing the disease compared with subjects in the bottom quartile. When we externally replicated these EPIC results using the independent MCCS cohort from Melbourne, Australia, we observed nearly identical associations.

The ATBC study recruited participants within a homogeneous population of Finnish men for randomized supplementation

with alpha-tocopherol, beta-carotene or placebo (23–25), whereas the EPIC study is a population-based observational study where recruitment took place across 10 Western European countries. The implications of these differences in study design and source populations are unclear, but may complicate direct comparisons between the studies (12). The biochemical measurements of vitamin B6 were also performed using different methodologies, ATBC applying the tyrosine decarboxylase assay and the EPIC and MCCS studies using liquid chromatography/tandem mass spectrometry. While both methods provide measurements of the active form of vitamin B6 (pyridoxal-5' phosphate), we note that ATBC controls had approximately 25% lower concentrations of vitamin B6 than did current smoking EPIC controls. It is unclear whether these differences explain the contrasting results, but several observations within the current study suggest that our findings were not because of chance or bias. For instance, the association between vitamin B6 concentrations and risk was nearly identical when comparing the cases with a separate control population, as well as when separately analyzing the independent cases and controls from the MCCS cohort. Taken together, these observations indicate that the association of vitamin B6 with risk is: 1) statistically robust and consistent across distinct European and Australian populations, 2) not because of random fluctuations in the case or control populations, and 3) not because of differential storage conditions or preanalytical treatment of cases and controls.

Several established risk factors for RCC could theoretically explain the association of vitamin B6 concentrations with risk, including tobacco smoking and obesity related factors, but the OR estimates were at most marginally affected when adjusting for these factors. Stratified analysis also showed that the decrease in

risk was evident up to 10 years after blood draw, suggesting that prediagnostic malignancies are unlikely to explain the results. We note that the association with risk was more prominent for smokers than never- and former smokers and for subjects without hypertension than for those with hypertension. While tobacco smoking is considered an established risk factor of RCC, it does not confer a particularly large increase in risk, current smokers having approximately a 50% risk increase compared with never smokers. That the effect estimate of the association between vitamin B6 and risk was twice that of being a current smoker would seem to rule out the possibility that tobacco exposure could be the underlying explanation for the observed association with vitamin B6. We were also able to control for circulating cotinine as a biomarker for recent tobacco exposure among current smokers. While adjustment for potential confounders did not affect the estimates substantially in either the EPIC or MCCS samples, we note that the set of potential confounders considered was the same in each study. As such, we cannot completely exclude the possibility that confounding by other unknown risk factors may partially account for the observed associations.

Our analysis also indicated that RCC cases with higher prediagnostic plasma B6 concentrations experience improved survival, the hazard ratio comparing the top and bottom quartiles being 0.57, an observation similar to the corresponding association with risk. While we mainly focused the survival analysis on all-cause mortality, an analysis of cause-specific mortality indicated that the association of B6 was driven by deaths caused by RCC. Indeed, no association between vitamin B6 and all-cause mortality was discernible in controls. Subsequent survival analyses among the cases from MCCS resulted in very similar hazard ratio estimates of all-cause mortality as in EPIC, thus providing an external replication that higher vitamin B6 was associated with improved survival. Similar to the EPIC analysis, the hazard ratio estimates were notably stronger in MCCS when considering cause-specific mortality of RCC.

We had limited information on stage at diagnosis for cases, being 45% complete in EPIC and 94% in MCCS, but adjusting for stage in these subsets provided similar and partly attenuated HR estimates. This result indicates that the survival benefit for subjects with higher vitamin B6 is at least partially mediated through stage. One interpretation of these findings is that lower B6 levels contribute to a greater risk of developing a more aggressive form of kidney cancer that presents with a later stage.

The current study was primarily initiated based on the hypothesis of one-carbon metabolism being important in renal cell carcinoma (9). However, multiple pathways implicated in cancer development are dependent on vitamin B6, examples of which include the tryptophan metabolism pathway, which is involved in immune function and inflammatory processes (26–28), as well as cell proliferation, angiogenesis, and oxidative stress (29), the latter being particularly relevant in smoking-related cancers. Our results should therefore not be taken as evidence for a causal role of dietary intake of vitamin B6 in RCC aetiology (30). Indeed, dietary intake of vitamin B6 is not strongly related to circulating vitamin B6 levels and has not been associated with RCC risk (31,32). This observation would seem to warrant further research on lifestyle and metabolic determinants of circulating vitamin concentrations. Indeed, given the number of enzymatic reactions that are dependent on vitamin

B6—over 100 having been identified to date (33,34)—adopting a more agnostic approach, such as metabolomics-based methodologies, may shed further light on the metabolic pathways involving vitamin B6 in the pathogenesis of renal cell carcinoma. Such an approach might also benefit from the use of multiple assessments per research participant in order to examine the extent to which within-person variability affects estimated associations between metabolites and risk.

High circulating concentrations of vitamin B6 were strongly associated with decreased risk of RCC. Our results also suggest a potentially important association with survival after RCC diagnosis. Further elucidating the metabolic pathways underlying this relationship and the extent to which it can be influenced by changes in lifestyle may suggest preventive strategies for RCC and other cancers, as well as improved cancer survival.

## References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer*. 2010;127(12):2893–2917.
2. Chow WH, Dong LM, Devesa SS. Epidemiology and risk factors for kidney cancer. *Nat Rev Urol*. 2010;7(5):245–257.
3. Scelo G, Brennan P. The epidemiology of bladder and kidney cancer. *Nat Clin Pract Urol*. 2007;4(4):205–217.
4. Purdue MP, Johansson M, Zelenika D, et al. Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nat Genet*. 2011;43(1):60–65.
5. Lee JE, Mannisto S, Spiegelman D, et al. Intakes of fruit, vegetables, and carotenoids and renal 492 cell cancer risk: a pooled analysis of 13 prospective studies. *Cancer Epidemiol Biomarkers*. 2009;18(6):1730–1739.
6. Wiseman M. The second World Cancer Research Fund/American Institute for Cancer Research expert report. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. *Proc Nutr Soc*. Aug 2008;67(3):253–256.
7. Buchner FL, Bueno-de-Mesquita HB, Linseisen J, et al. Fruits and vegetables consumption and the risk of histological subtypes of lung cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Cancer Causes Control*. 2010;21(3):357–371.
8. Buchner FL, Bueno-de-Mesquita HB, Ros MM, et al. Variety in fruit and vegetable consumption and the risk of lung cancer in the European prospective investigation into cancer and nutrition. *Cancer Epidemiol Biomarkers Prev*. 2010;19(9):2278–2286.
9. Kim YI. Folate and colorectal cancer: an evidence-based critical review. *Mol Nutr Food Res*. 2007;51(3):267–292.
10. Lee JE, Mannisto S, Spiegelman D, et al. Intakes of fruit, vegetables, and carotenoids and renal cell cancer risk: a pooled analysis of 13 prospective studies. *Cancer Epidemiol Biomarkers Prev*. 2009;18(6):1730–1739.
11. Gibson TM, Weinstein SJ, Mayne ST, et al. A prospective study of one-carbon metabolism biomarkers and risk of renal cell carcinoma. *Cancer Causes Control*. 2010;21(7):1061–1069.
12. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr*. 2002;5(6B):1113–1124.
13. Giles GG, English DR. The Melbourne Collaborative Cohort Study. *LARC Sci Publ*. 2002;156:69–70.
14. Muller DC, Baglietto L, Manning JT, et al. Second to fourth digit ratio (2D:4D), breast cancer risk factors, and breast cancer risk: a prospective cohort study. *Br J Cancer*. 2012;107(9):1631–1636.
15. Midttun O, Hustad S, Ueland PM. Quantitative profiling of biomarkers related to B-vitamin status, tryptophan metabolism and inflammation in human plasma by liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom*. 2009;23(9):1371–1379.

16. Ueland PM, Midttun O, Windelberg A, Svardal A, Skalevik R, Hustad S. Quantitative profiling of folate and one-carbon metabolism in large-scale epidemiological studies by mass spectrometry. *Clin Chem Lab Med*. 2007;45(12):1737–1745.
17. Midttun O, Kvalheim G, Ueland PM. High-throughput, low-volume, multianalyte quantification of plasma metabolites related to one-carbon metabolism using HPLC-MS/MS. *Anal Bioanal Chem*. 2013;405(6):2009–2017.
18. Molloy AM, Scott JM. Microbiological assay for serum, plasma, and red cell folate using cryopreserved, microtiter plate method. *Methods Enzymol*. 1997;281:43–53.
19. Kelleher BP, Broin SD. Microbiological assay for vitamin B12 performed in 96-well microtitre plates. *J Clin Pathol*. 1991;44(7):592–595.
20. Bernards CA, Belin TR, Schafer JL. Robustness of a multivariable normal approximation for imputation of incomplete binary data. *Stat Med*. 2007;26(6):1368–1382.
21. Lee KJ, Carlin JB. Multiple Imputation for Missing Data: Fully Conditional Specification Versus Multivariable Normal Imputation. *Am J Epidemiol*. 2010;171(5):624–632.
22. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med*. 2002;21(15):2175–2197.
23. The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. The ATBC Cancer Prevention Study Group. *Ann Epidemiol*. 1994;4(1):1–10.
24. Lao O, Lu TT, Nothnagel M, et al. Correlation between genetic and geographic structure in Europe. *Curr Biol*. 2008;18(6):1241–1248.
25. Palo JU, Ulmanen I, Lukka M, Ellonen P, Sajantila A. Genetic markers and population history: Finland revisited. *Eur J Hum Genet*. 2009;17(10):1336–13346.
26. Midttun O, Ulvik A, Ringdal Pedersen E, et al. Low plasma vitamin B-6 status affects metabolism through the kynurenine pathway in cardiovascular patients with systemic inflammation. *J Nutr*. 2011;141(4):611–617.
27. Platten M, Wick W, Van den Eynde BJ. Tryptophan catabolism in cancer: beyond IDO and tryptophan depletion. *Cancer Res*. 2012;72(21):5435–5440.
28. Sakakeeny L, Roubenoff R, Obin M, et al. Plasma pyridoxal-5-phosphate is inversely associated with systemic markers of inflammation in a population of U.S. adults. *J Nutr*. 2012;142(7):1280–1285.
29. Matsubara K, Komatsu S, Oka T, Kato N. Vitamin B6-mediated suppression of colon tumorigenesis, cell proliferation, and angiogenesis (review). *J Nutr Biochem*. 2003;14(5):246–250.
30. Giovannucci E. Nutrient biomarkers are not always simple markers of nutrient intake. *Am J Clin Nutr*. 2013;97(3):657–659.
31. Cho E, Giovannucci EL, Joh HK. Nutrients related to one-carbon metabolism and risk of renal cell cancer. *Cancer Causes Control*. 2013;24(2):373–382.
32. Bosetti C, Scotti L, Maso LD, et al. Micronutrients and the risk of renal cell cancer: a case-control study from Italy. *Cancer Causes Control*. 2007;120(4):892–896.
33. Wu XY, Lu L. Vitamin b6 deficiency, genome instability and cancer. *Asian Pac J Cancer Prev*. 2012;13(11):5333–5338.
34. Bairoch A. The ENZYME database in 2000. *Nucleic Acids Res*. 2000;28(1):304–305.

## Funding

The World Cancer Research Fund (UK) supported the biochemical analyses for the European Prospective Investigation into Cancer and Nutrition (EPIC) study. The funding organization had no role in design and conduct of the study, collection, management, analysis, or interpretation of the data, preparation, review, or approval of the manuscript. The work undertaken by DCM that was reported in this paper was done during the tenure of an International Agency for Research on Cancer (IARC)–Australia Postdoctoral Fellowship from the International Agency for Research on Cancer, supported by the Cancer Council Australia. Dr. Ueland reports that he is a member of the steering board of the nonprofit Foundation to Promote Research Into Functional Vitamin B12 Deficiency.

The EPIC study has been supported by the Europe Against Cancer Program of the European Commission (SANCO); Deutsche Krebshilfe, Deutsches Krebsforschungszentrum, German Federal Ministry of Education and Research;

Danish Cancer Society; Health Research Fund (FIS) of the Spanish Ministry of Health, Spanish Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra; Catalan Institute of Oncology, Spain; the ISCIII of the Spanish Ministry of Health (RETICC DR06/0020); Cancer Research UK; Medical Research Council, UK; Greek Ministry of Health; Stavros Niarchos Foundation; Hellenic Health Foundation; Italian Association for Research on Cancer (AIRC); Italian National Research Council, Fondazione-Istituto Banco Napoli, Italy; Associazione Italiana per la Ricerca sul Cancro-AIRC-Milan; Compagnia di San Paolo; Dutch Ministry of Public Health, Welfare and Sports; World Cancer Research Fund; Swedish Cancer Society; Swedish Scientific Council; Regional Government of Västerbotten, Sweden; NordForsk (Centre of excellence programme HELGA), Norway; French League against Cancer (LNCC), France; National Institute for Health and Medical Research (INSERM), France; Mutuelle Générale de l'Éducation Nationale (MGEN), France; 3M Co, France; Gustave Roussy Institute (IGR), France; and General Councils of France.

## Notes

Author contributions: MJ and PB initiated, acquired the main funding, and designed this investigation, with additional funding for the Melbourne Collaborative Cohort Study (MCCS) analysis provided by GS, PMU, SEV, and OM led the laboratory analysis. MJ led the statistical analysis of EPIC data with important contributions from AF, DCM, JKB and DCM led the statistical analysis of MCCS data. MJ drafted the first version of the manuscript with important contributions from AF, DCM, JKB, PMU, GS, and PB. All authors were involved with collection of data, data interpretation, critical revision of the manuscript, and approval of the final version. MJ and PB had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. ER is the overall coordinator of the EPIC study, which he designed and implemented in collaboration with the main investigators in the collaborating centers.

We declare that we have no conflicts of interest.

**Affiliations of authors:** International Agency for Research on Cancer, Lyon, France (MJ, AF, DCM, PB); Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, Australia (JKB, GG, GiS); Bevilacqua AS, Bergen, Norway (OM); Norwegian Institute of Public Health and Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway (SEV); Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK (RCT); Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute (ISPO), Florence, Italy (DP); Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy (AM); Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy (SS); Hellenic Health Foundation, Athens, Greece (AT, DT); World Health Organization Collaborating Center for Food and Nutrition Policies, Department of Hygiene, Epidemiology and Medical Statistics, University of Athens Medical School, Athens, Greece (AT, PL); Department of Epidemiology, Harvard School of Public Health, Boston, MA (PL, DT); Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece (PL, DT); Department of Surgical and Perioperative Sciences, Urology and Andrology, Umeå University, Umeå, Sweden (BL); Department of Biobank Research, Umeå University, Umeå, Sweden (GH); Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, Tromsø, Norway (EW, GuS); Department of Research, Cancer Registry of Norway, Oslo, Norway (EW); Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden (EW); Samfundet Folkhälsan, Helsinki, Finland (EW); Catalan Institute of Oncology (ICO), Barcelona, Spain (CAG); Public Health Direction, Basque regional Health Department, San Sebastian, Spain (MD); Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP) - Health Research Institute (BIODONOSTIA), Spain (HBBdM, MMR); Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, the Netherlands (PHP); National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands (HBBdM, MMR); Department of Gastroenterology and Hepatology, University Medical Centre, Utrecht, the Netherlands (HBBdM); Inserm, Centre for research in

Epidemiology and Population Health (CESP), U1018, Nutrition, Hormones and Women's Health team, Villejuif, France (MCBR, GF, FC); Université Paris Sud, UMRS 1018, Villejuif, France (MCBR, GF, FC); Institut Gustave-Roussy (IGR), Villejuif, France (MCBR, GF, FC); Andalusian School of Public Health, Granada, Spain (MJS); Consortium for Biomedical Research in Epidemiology and Public Health (CIBER Epidemiología y Salud Pública-CIBERESP), Spain (MJS, ABG, CN); Navarra Public Health Institute, Pamplona, Spain (ABG); Department of Epidemiology, Regional Health Authority, Murcia, Spain (CN); Unit of Preventive Medicine and Public Health, School of Medicine, University of Murcia, Spain (CN); Public Health Directorate, Asturias, Spain (JRQ); Section for Epidemiology, Department of Public Health, Aarhus University, Aarhus, Denmark (KO); Danish Cancer Society Research Center, Copenhagen, Denmark (AT); Department of Epidemiology, German Institute of Human Nutrition (DIfE) Potsdam-Rehbrücke, Nuthetal, Germany (HB, KA); Department of Epidemiology and Biostatistics, School of Public Health, Imperial College, London, UK (PV, MJG, ER); HuGeF Foundation, Torino, Italy (PV, GiS); Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany (RK); Institute of Genetic Medicine, Newcastle University, Newcastle, UK (CR); Section for Pharmacology, Institute of Medicine, University of Bergen, Bergen, Norway (PMU); Laboratory of Clinical Biochemistry, Haukeland University Hospital, Bergen, Norway (PMU); Centre for Epidemiology and Biostatistics, University of Melbourne, Australia (GG, GiS).

# **CHAPTER V:**

***B-VITAMINS AND LUNG***

***CANCER IN THE LC3***

## Objective

In addition to investigate whether OCM biomarkers were associated with other smoking-related cancers than lung cancer, we wanted to externally validate and support some previous findings from the EPIC lung cancer study<sup>13</sup>. We then further investigated the association between circulating concentrations of vitamin B6, folate and methionine with lung cancer risk in the LC3 study which gather multiple populations from America, Europe, Asia and Australia.

## Selection of cases and controls

In total, twenty nested case-control studies contributed in this study, gathering over 5,500 case-control pairs. Further study design information of each participated cohort is described in Appendix 1. Lung cancer cases were defined on the basis of the International Classification of Disease for Oncology, Second Edition (ICD-O-2), and included invasive cancers coded as C34.0-C34.9. After excluding cases who were not correctly matched on smoking status (n=124 cases), who had insufficient plasma samples (n=42), or had a revised date of diagnosis prior to blood draw (n=13), 5,364 lung cancer cases and their matched controls remained eligible.

For each case, one control was randomly chosen from risk-sets consisting of all cohort members alive and free of cancer (except non-melanoma skin cancer) at the time of diagnosis of the index case. Matching criteria were cohort, sex, date of blood collection ( $\pm 1$  month, relaxed to  $\pm 3$  months for sets without available controls), and date of birth ( $\pm 1$  year, relaxed to  $\pm 3$  years). To adequately control for confounding by tobacco exposure controls were also matched for 3 categories of smoking status including never smoker, short/long term quitters among former smokers (< 10 years/ 10+ years since quitting), and light/heavy smokers among current smokers (< 15 years/ 15+ cigarettes per day).

## Data harmonization

The LC3 consortium was led by our group, the Genetic Epidemiology Group, at the International Agency for Research on Cancer. Once the consortium started, I was strongly involved in the coordination of biosample shipments. We created a data dictionary (see Appendix 2) which was sent to all participated cohorts including epidemiological and clinical

variables. With support from Dr Brennan and Dr Johansson, I mostly coordinated and led the data management of the LC3 study database by completing the study database including biochemical, epidemiological and clinical data for all participating cohorts, and harmonizing the baseline data. The LC3 database includes 44 biomarkers and over 200 variables for over 12,000 participants.

### **Biochemical analyses**

All biosamples from the LC3 study were also sent on dry ice to the Bevital A/S laboratory, in Bergen, Norway, which permit us to have a good comparability with the two previous studies (Chapter III and IV). Additional biochemical analysis information has been provided previously in Chapter II (see *Material and Methods, Biochemical analyses*).

### **Statistical analyses**

In the LC3 study we used conditional logistic regression in order to investigate circulating OCM biomarkers in relation to lung cancer risk. In addition, hierarchical linear models were used to describe within- and between-cohort differences in (log) concentrations of each biomarker by demographic and host factors. These calculations were performed using SAS 9.2, R version 3.1.3 and Stan version 2.6.0. Supplementary statistical details are provided in Chapter II (see *Material and Methods, Statistical methods*).

# Article 3

## Circulating Folate, Vitamin B6 and Methionine in relation to Lung Cancer Risk in the Lung Cancer Cohort Consortium

Authors: **Anouar Fanidi\***, David Muller\*, Jian-Min Yuan\*, Victoria L. Stevens\*, Stephanie J. Weinstein\*, Demetrius Albanes\*, Ross Prentice\*, Cynthia A. Thomsen, Mary Pettinger, Qiuyin Cai, William J. Blot, Jie Wu, Alan A. Arslan, Anne Zeleniuch-Jacquotte,, Marjorie L. McCullough, Loic Le Marchand, Lynne R. Wilkens, Christopher A. Haiman, Xuehong Zhang, Jiali Han, Meir J. Stampfer, Stephanie A. Smith-Warner, Edward Giovannucci, Graham G. Giles, Allison M Hodge, Gianluca Severi, Mikael Johansson, Kjell Grankvist, Arnulf Langhammer, Steinar Krokstad, Marit Næss, Renwei Wang, Yu-Tang Gao, Lesley M. Butler, Woon-Puay Koh, Xiao-Ou Shu, Yong-Bing Xiang, Honglan Li, Wei Zheng, Qing Lan, Kala Visvanathan, Judith Hoffman Bolton, Per Magne Ueland, Øivind Midttun, Arve Ulvik, Neil Caporaso, Mark Purdue, Regina Zeigler, Neal D. Freedman, Julie E. Buring, I-Min Lee, Howard D. Sesso, J. Michael Gaziano, Jonas Manjer, Ulrika Ericson, Caroline Relton, Paul Brennan\*\*, Mattias Johansson\*\*

*Submitted to Journal of National Cancer Institute*



## Circulating Folate, Vitamin B6 and Methionine in relation to Lung Cancer Risk in the Lung Cancer Cohort Consortium (LC3)

Anouar Fanidi<sup>1\*</sup>, David Muller<sup>1,2\*</sup>, Jian-Min Yuan<sup>3,4\*</sup>, Victoria L. Stevens<sup>5\*</sup>, Stephanie J. Weinstein<sup>6\*</sup>, Demetrius Albanes<sup>6\*</sup>, Ross Prentice<sup>7\*</sup>, Cynthia A. Thomsen<sup>8</sup>, Mary Pettinger<sup>7</sup>, Qiuyin Cai<sup>9</sup>, William J. Blot<sup>9,10</sup>, Jie Wu<sup>9</sup>, Alan A. Arslan<sup>11</sup>, Anne Zeleniuch-Jacquotte<sup>12</sup>, Marjorie L. McCullough<sup>5</sup>, Loic Le Marchand<sup>13</sup>, Lynne R. Wilkens<sup>13</sup>, Christopher A. Haiman<sup>13</sup>, Xuehong Zhang<sup>14</sup>, Jiali Han<sup>15</sup>, Meir J. Stampfer<sup>14,15,16</sup>, Stephanie A. Smith-Warner<sup>15,16</sup>, Edward Giovannucci<sup>14,15,16</sup>, Graham G. Giles<sup>17,18</sup>, Allison M Hodge<sup>17</sup>, Gianluca Severi<sup>17,19,20,21</sup>, Mikael Johansson<sup>22</sup>, Kjell Grankvist<sup>22</sup>, Arnulf Langhammer<sup>23</sup>, Steinar Krokstad<sup>23</sup>, Marit Næss<sup>23</sup>, Renwei Wang<sup>3</sup>, Yu-Tang Gao<sup>24</sup>, Lesley M. Butler<sup>3,4</sup>, Woon-Puay Koh<sup>25</sup>, Xiao-Ou Shu<sup>26</sup>, Yong-Bing Xiang<sup>27</sup>, Honglan Li<sup>27</sup>, Wei Zheng<sup>26</sup>, Qing Lan<sup>27</sup>, Kala Visvanathan<sup>28</sup>, Judith Hoffman Bolton<sup>28</sup>, Per Magne Ueland<sup>29,30</sup>, Øivind Midttun<sup>31</sup>, Arve Ulvik<sup>31</sup>, Neil Caporaso<sup>6</sup>, Mark Purdue<sup>6</sup>, Regina Zeigler<sup>6</sup>, Neal D. Freedman<sup>6</sup>, Julie E. Buring<sup>32,33,34</sup>, I-Min Lee<sup>32,34</sup>, Howard D. Sesso<sup>32,33,34</sup>, J. Michael Gaziano<sup>32,33,35</sup>, Jonas Manjer<sup>36</sup>, Ulrika Ericson<sup>37</sup>, Caroline Relton<sup>38,39</sup>, Paul Brennan<sup>1\*\*</sup>, Mattias Johansson<sup>1\*\*</sup>

\*Co-first authors

\*\*These authors jointly supervised this work

Corresponding author:

Paul Brennan

International Agency for Research on Cancer (IARC/WHO)

150 cours Albert Thomas

69372 Lyon cedex 08, France

Phone: +33 (0) 4 72 73 83 91

Fax: +33 (0) 4 72 73 83 42

Email: gep@iarc.fr

Affiliations:

1. Genetic Epidemiology Group, International Agency for Research on Cancer, Lyon, France
2. Department of Epidemiology and Biostatistics, Imperial College London
3. Division of Cancer Control and Population Sciences, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania, USA.
4. Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
5. Epidemiology Research Program, American Cancer Society, Inc. 250 Williams St. Atlanta, GA 30303
6. Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH
7. Division of Public Health Sciences Fred Hutchinson Cancer Research Center 1100 Fairview Ave. N, M3-A410 Seattle, Washington 98109, U.S.A.
8. Health Promotion Sciences, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, Arizona
9. Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA
10. Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA. International Epidemiology Institute, Rockville, MD, USA
11. Departments of Obstetrics and Gynecology, Population Health and Environmental Medicine, New York University School of Medicine
12. Department of Population Health, New York University School of Medicine
13. Epidemiology program, Cancer Research Center of Hawaii, University of Hawaii
14. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.
15. Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA.
16. Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA.
17. Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Australia.

18. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Victoria, Australia
19. Molecular end Epidemiology Unit, HuGeF, Human Genetics Foundation, Torino, Italy.
20. Inserm (Institut National de la Sante et de la Recherche Medicale), Centre for Research in Epidemiology and Population Health, Villejuif 94805, France.
21. Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Australia
22. Umeå, University, Umeå, Sweden
23. HUNT Research Centre, Department of Public Health and General Practice, NTNU, Norwegian University of Science and Technology, Levanger, Norway
24. Department of Epidemiology, Shanghai Cancer Institute, Shanghai Jiaotong University, Shanghai, China
25. Duke-NUS Graduate Medical School Singapore, Singapore
26. Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA
27. Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
28. Johns Hopkins University Bloomberg School of Public Health Department of Epidemiology George W Comstock Center for Public Health Research and Prevention Health Monitoring Unit, USA
29. Department of Clinical Science, University of Bergen, Bergen, Norway Laboratory of Clinical Biochemistry,
30. Haukeland University Hospital, Bergen, Norway
31. Bevital AS, 5021 Bergen, Norway
32. Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA USA.
33. Division of Aging, Department of Medicine, Brigham and Women's Hospital, Boston, MA USA.
34. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA USA
35. Boston VA Medical Center, Boston, MA USA

36. Department of Surgery, Skåne University Hospital Malmö Lund University, Malmö Sweden
37. Department of clinical sciences Malmö, Lund University, Sweden
38. Institute of Genetic Medicine, Newcastle University, Newcastle, United Kingdom
39. MRC Integrative Epidemiology Unit, School of Social & Community Medicine, University of Bristol, Bristol, United Kingdom

Keywords: one-carbon metabolism, lung cancer, consortium, prospective study

Abbreviations:

CI confidence interval

EPIC European Prospective Investigation into Cancer and Nutrition

OCM one-carbon metabolism

OR odds ratio

LC3 Lung Cancer Cohort Consortium

Article category: Research article - Epidemiology

Number of words - article: 2,998

Number of Tables: 4

Number of Figures: 1

Number of Supplementary Tables: 4

Number of Supplementary Figures: 14

Number of References: 24

## Abstract

**Background** – Circulating concentrations of B-vitamins and factors related to one-carbon metabolism have been found to be strongly inversely associated with lung cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. The extent to which these associations are present in other study populations is unknown.

**Methods** – Within 20 prospective cohorts from the NCI Cohort Consortium, a nested case-control study was designed including 5,364 incident lung cancer cases and 5,364 controls that were individually matched to cases by age, sex, cohort, and smoking status. Centralized biochemical analyses were performed to measure circulating concentrations of vitamin B6, folate and methionine, as well as cotinine as an indicator of recent tobacco exposure. The association between these biomarkers and lung cancer risk was evaluated using conditional logistic regression models.

**Results** – Participants with higher circulating concentrations of vitamin B6 and folate had a modestly decreased risk of lung cancer risk overall, the odds ratio when comparing the top and bottom fourths [OR<sub>4vs1</sub>] being 0.88 (95% confidence interval [95% CI], 0.78-1.00) and 0.86 (95% CI 0.74-0.99), respectively. We found stronger associations among for men (OR<sub>4vs1</sub> 0.74; 95% CI 0.62-0.89 for vitamin B6, OR<sub>4vs1</sub> 0.75; 95% CI 0.61-0.93 for folate) and ever smokers (OR<sub>4vs1</sub> 0.79; 95% CI 0.68-0.91 for vitamin B6, OR<sub>4vs1</sub> 0.85; 95% CI 0.72-0.99 for folate). We further noted that the association of folate was restricted to Europe/Australia and Asia, whilst no clear association was observed for the United-States. Circulating concentrations of methionine were not associated with lung cancer risk overall or in important subgroups.

**Conclusion** – Although confounding by tobacco exposure or reverse causation cannot be ruled out, these study results are consistent with a small decrease in lung cancer risk in ever smokers who avoid low concentrations of circulating folate and vitamin B6.

## Introduction

The most important and effective means for lung cancer control is by reducing the number of people who smoke tobacco products. However, even among subjects who have quit smoking, the lifetime risk of lung cancer remains high,<sup>249-251</sup> and in some regions where anti-tobacco campaigns have been successful, such as the US, a large proportion of lung cancer cases now occurs among former and never smokers. Given these considerations, exploring additional means of primary prevention of lung cancer is important for subgroups of high risk—in particular former smokers—who seek additional means to further reduce their risk.

One-carbon metabolism (OCM) encompasses a series of biochemical reactions involving B-vitamins that are essential to ensure balanced DNA synthesis and methylation,<sup>133 180</sup> and changes in specific OCM factors have been implicated in cancer development.<sup>132 180 195</sup> If imbalances in B-vitamins such as folate (B9) and vitamin B6 are proved to be causally implicated in lung cancer etiology, they would provide an appealing target for chemoprevention as they are modifiable by changes in diet or supplementation.<sup>145 252</sup>

Only two prospective studies have been published on lung cancer and circulating biomarkers of OCM, the EPIC study<sup>13</sup> and the Alpha-Tocopherol, Beta-Carotene, Prevention Cancer (ATBC) study of smoking men,<sup>185</sup> both of which reported an inverse association between circulating vitamin B6 and lung cancer risk. In particular, the EPIC nested case-control study of 899 cases and 1,770 controls reported that the inverse association of vitamin B6 with lung cancer risk was strong and consistent regardless of tobacco smoking history, and also reported strong inverse associations of methionine and folate with risk.<sup>13</sup>

In order to provide conclusive evidence on the importance of circulating B-vitamins and lung cancer risk, we initiated the Lung Cancer Cohort Consortium (LC3) with a combined cohort population of over 2,000,000 participants from North America, Europe, Asian, and Australia, to retrieve blood samples and conduct biochemical analysis on over 5,000 case-controls pairs.

## **Material and Methods**

### **Study population**

We invited all prospective cohort studies with cryopreserved baseline plasma and serum samples that in 2009 were members in the US National Cancer Institute (NCI) Cohort Consortium to participate in the study. Twenty cohorts fulfilled those criteria and accepted to participate, resulting in a combined cohort population of over 2,000,000 participants from North America, Europe, Asia and Australia. Brief details on design of the cohorts and their follow-up procedures are provided in the Supplementary methods.

### **Selection of cases and controls**

Lung cancer cases were defined on the basis of the International Classification of Diseases for Oncology, Second Edition (ICD-O-2), and included all invasive cancers coded as C34.0-C34.9. We selected a total 5,545 lung cancer cases, and in order optimize the statistical power in smoking stratified analyses, never and former smoking cases were oversampled. For each case, one control was matched by cohort, sex, date of blood collection ( $\pm 1$  month, relaxed to  $\pm 3$  months for sets without available controls), and date of birth ( $\pm 1$  year, relaxed to  $\pm 3$  years), as well as smoking status in 5 categories; never smokers, short and long term quitters among former smokers ( $<10$  years,  $\geq 10$  years since quitting), and light and heavy smokers among current smokers ( $< 15$  years,  $\geq 15$  cigarettes per day). After various exclusions (see Supplementary methods), 5,364 lung cancer case-control pairs remained eligible for the risk analysis.

### **Biochemical analyses**

Concentrations of vitamin B6, methionine and cotinine were determined by mass spectrometry based methods (LC-MS/MS, GC-MS/MS),<sup>239 240</sup> and microbiological methods were used to determine concentrations of folate (*Lactobacillus casei*).<sup>241</sup> Further details are provided in Supplementary methods.

### **Statistical analyses**

Hierarchical linear models were used to describe the variation in average biomarker concentrations between the cohorts, and the extent to which these could be explained by differences in baseline characteristics. Relative risks of lung cancer were estimated by

calculating odds ratios (OR) and 95% confidence intervals (CI) using conditional logistic regression (conditioning on matched case-control sets) with the first quartile as the referent. Additional covariates were included to account for confounding by risk factors, including indicators of educational attainment (in six categories) and tobacco exposure (in addition to matching to smoking status by design: cotinine concentrations [quartiles defined in current smokers]). Including additional covariates of body mass index (BMI) and alcohol intake did not appreciably alter the results and were not included in the final models. As a sensitivity analysis, we fitted models that were additionally adjusted for smoking duration or pack-years of smoking among ever smokers. All risk analyses were conducted overall, and stratified by smoking status and region (US, EU/AU, Asia).

As an indication of the overall statistical strength of association between each biomarker and risk, we calculated a p-value for trend by including the base-2 logarithm ( $\log_2$ ) of the biomarker concentration as a continuous variable in a separate conditional logistic regression model. The same approach was used in stratified risk analyses according to other pre-defined demographic characteristics and risk factors. OR estimates per  $\log_2$  unit ( $\log_2$  OR) may be interpreted as the relative risk associated with a doubling in the concentration of a circulating biomarker.

All statistical analyses were conducted using SAS 9.2 (Cary, North Carolina),<sup>245</sup> R version 3.1.3<sup>247</sup> or Stan version 2.9.0.<sup>248</sup>

Additional details on the statistical methods are provided in the Supplementary methods.

## Results

### Baseline Characteristics

Baseline characteristics of the study subjects separated by geographic areas are shown in Table 1. The final study population included 5,364 lung cancer cases and 5,364 individually matched controls. Median age at cohort enrolment was 63 years and the median time between blood draw and lung cancer diagnosis was 6.3 years. Because we oversampled never and former smokers, approximately half of the study population were current smokers at recruitment (2,519 case-control pairs, 47%), whereas the other half were either former (1,518 case-control pairs, 28%) or never smokers (1,327 case-control pairs, 25%).



Overall, 46% of the participants were women, though the sex distribution varied by region, the US cohorts contributing more women and the cohorts from Europe and Asia contributed more men.

### **Variation in circulating biomarkers**

There was substantial between-cohort variation in concentrations of all biomarkers (Table 1), in particular for folate and vitamin B6 concentrations that were substantially higher in US cohorts than for EU/AU or Asian cohorts. This could only partly be explained by differences in age, sex, smoking status, or BMI (Figure 1). For circulating vitamin B6 and methionine, the between-cohort variability was similar before and after covariate adjustment (Figure 1). Conversely, for circulating folate, adjustment for age, sex, smoking status, and BMI reduced the between-cohort variability, the sd being 0.70 (90% CI 0.54-0.97) without adjustment and 0.44 (90% CI 0.32-0.65) with adjustment.

### **Risk analysis overall and stratified by region**

Participants with higher circulating vitamin B6 and folate had an approximate 20% lower risk of lung cancer overall (Table 2 and 3), the odds ratio when comparing the top and bottom quartiles [ $OR_{4vs1}$ ] being 0.81 for vitamin B6 (95% CI 0.72-0.92) and 0.80 for folate (95% CI 0.70-0.92). These OR estimates were slightly attenuated after adjustment for circulating cotinine and education ( $OR_{4vs1}$  0.88, 95% CI 0.78-1.00 for vitamin B6 and  $OR_{4vs1}$  0.86, 95% CI 0.74-0.99 for folate). We did not observe any clear association between circulating methionine and risk of lung cancer, the overall  $OR_{4vs1}$  being 0.95 (95%, 0.85-1.07, Table 4). Among participants with available information on smoking duration and pack-years of smoking (90% of ever smokers), further adjustment by continuous number of years of smoking or pack-years of smoking did not notably affect the estimates (Supplementary Table 1-4).

In risk analyses stratified by region (Table 2-4), the overall inverse associations of vitamin B6 and folate were mainly driven by Asian and EU/AU participants. The adjusted  $OR_{4vs1}$  for vitamin B6 was 0.82 in the Asian cohorts (95% CI 0.65-1.02) and 0.78 for the EU/AU cohorts (95% CI 0.62-1.00), and 0.95 for the US cohorts (95% CI 0.80-1.13). Similarly, for folate, the

adjusted  $OR_{4vs1}$  was 0.84 for the Asian cohorts (95% CI 0.68-1.03) and 0.77 for the EU/AU cohorts (95% CI 0.60-0.99), and 0.95 for the US cohorts (95% CI 0.78-1.15).

### **Risk analysis stratified by smoking status and sex**

In analyses stratified by smoking status, we observed an inverse association between vitamin B6 and overall risk for former ( $OR_{4vs1}$  0.72, 95% CI 0.57-0.91) and current smokers ( $OR_{4vs1}$  0.79, 95% CI 0.65-0.95) (Table 2). Folate was also inversely associated with risk for former ( $OR_{4vs1}$  0.66, 95% CI 0.51-0.85), but not for current smokers ( $OR_{4vs1}$  0.97, 95% CI 0.77-1.21) (Table 3). When analyzing ever smokers overall (i.e. former and current smokers combined), we observed an inverse association for circulating vitamin B6 ( $OR_{4vs1}$  0.79, 95% CI 0.68-0.91) and folate ( $OR_{4vs1}$  0.85, 95% CI 0.72-0.99). Of note, whilst no clear association was seen between folate and risk for never smokers ( $OR_{4vs1}$  0.86, 95% CI 0.63-1.17), we observed a positive association between vitamin B6 and risk for never smokers, the  $OR_{4vs1}$  being 1.51 (95% CI 1.14-2.01) (Table 2).

Further analyses stratified by sex showed that any inverse association between folate or vitamin B6 and lung cancer risk was restricted to ever smokers and men. First, the association of vitamin B6 in men ( $OR_{4vs1}$  0.74, 95% CI 0.62-0.89) was consistently observed in cohorts from each region (Table 2), with similar observations for folate (overall  $OR_{4vs1}$  0.75; 95% CI 0.61-0.93) (Table 3). Conversely, no association was seen among for women overall, nor for women separated by region (Table 2-3).

### **Stratified risk analysis by demographic and diagnostic parameters**

To assess potential of effect modification, we fitted models that included interaction terms between circulating folate or vitamin B6 and various covariates (Supplementary Figures 1-8). Gender modified the associations between circulating concentrations of vitamin B6 and folate and lung cancer risk ( $P_{heterogeneity} \leq 0.05$ , Supplementary Tables 1 and 4). When assessing blood samples taken up to 10 years prior to diagnosis, we observed that the overall inverse association between vitamin B6 and lung cancer risk was driven by cases diagnosed closer to blood draw ( $P_{heterogeneity} = 0.001$ , Supplementary Figure 1), as well as by former ( $OR_{log2}$  0.81; 95% CI 0.70-0.90, Supplementary Figure 3) and current smokers ( $OR_{log2}$  0.79; 95% CI 0.69-0.91, Supplementary Figure 4). Stratified analyses by histology showed a

stronger inverse association between vitamin B6 and folate and risk of lung squamous cell cancer ( $OR_{\log_2}$  0.86; 95% CI 0.77-0.95 for vitamin B6 and  $OR_{\log_2}$  0.88; 95% CI 0.78-0.98 for folate) than for other histological types (Supplementary Figures 1 and 5). Stratified analysis by period of blood collection before and after 1996 when folate fortification of food items took effect in US did not reveal any discernable difference in lung cancer risk for circulating folate (Supplementary Figure 12). The associations between biomarker concentrations and lung cancer risk for individual cohorts are presented in Supplementary Figures 9-11. We also estimated the potential risk increase associated with being clinically deficient in vitamin B6 and folate, and observed a 23% increase in risk of lung cancer for those deficient in vitamin B6 ( $OR = 1.23$ , 95% CI 1.09-1.38) that was more apparent for being diagnosed with squamous cell cancer ( $OR=1.54$ , 95% CI 1.18-2.02), and also those for developing lung cancer within the first 3 years following blood collection ( $OR=1.83$ , 95% CI 1.42-2.36) (Supplementary Figure 13). Similarly, for folate we observed an increased risk for those classified as deficient ( $OR=1.27$ , 95% CI 1.07-1.49, Supplementary Figure 14). Further stratified analyses are presented in Supplementary Figures 13 and 14.

## **Discussion**

Our aim was to evaluate the associations between lung cancer risk and circulating concentrations of folate, vitamin B6, and methionine, within a consortium of 20 prospective cohorts from the US, Europe, Asia and Australia. We found that individuals with low concentrations of vitamin B6 and folate have a small risk increase of lung cancer, particularly among men, current and former smokers, and participants living outside the United States.

### **Previous studies and the Lung Cancer Cohort Consortium (LC3)**

To date, this is the largest cohort consortium initiative for which circulating biomarkers were assessed in relation to a single cancer outcome with coordinated biochemical analysis conducted in one centralized laboratory. The background to this study was the initial findings from the EPIC cohort involving 899 lung cancer cases and 1,770 controls.<sup>13</sup> The EPIC study reported a strong inverse association between vitamin B6 and lung cancer regardless of smoking history, and additionally, found similar inverse associations with risk for folate and methionine. The results of the EPIC study were noteworthy as they suggested

a potentially important role of the one-carbon metabolism pathway in lung cancer etiology, in addition to tobacco exposure. Along with additional evidence from the ATBC cohort,<sup>185</sup> this motivated the organization of the Lung Cancer Cohort Consortium (LC3) in order to conclusively evaluate the extent to which these findings translate into other populations from Asia, the United States, Europe and Australia.

### **Interpreting the results**

The current study discerned relatively weak inverse associations between folate, vitamin B6 and lung cancer risk overall, associations that differed substantially between groups of different tobacco exposure history. For former and current smokers, we observed slightly lower risk among for those study participants with higher concentrations of folate and vitamin B6 compared to those having lower concentrations. This inverse association for lung cancer risk with vitamin B6 or folate can also be interpreted as an increased risk among participants in the lowest category of concentrations compared to those in the remaining three higher categories (Tables 2 and 3). In contrast, methionine was not clearly associated with risk overall or in any important subgroup.

Previous studies have shown that being a current smoker is clearly associated with lower circulating concentrations of folate and vitamin B6,<sup>13</sup> and our analysis confirms this relation (data not shown). This highlights the importance of carefully accounting for tobacco exposure in risk analyses of B-vitamins and lung cancer. We used circulating cotinine, an objective and accurate measure of current tobacco exposure,<sup>253</sup> and the weak inverse associations of vitamin B6 and folate with lung cancer risk remained after adjusting for cotinine. Further accounting for smoking duration and intensity did not substantively alter the OR estimates. In contrast to current smokers, former smokers tend to have similar circulating B-vitamin concentrations as never smokers.<sup>13</sup> This suggests that the potential confounding effect of tobacco exposure on the associations between circulating vitamin B6/folate and lung cancer risk is of less concern for former smokers than for current smokers. Despite these considerations, and given the important impact of both past and current tobacco exposure on lung cancer risk, our results do not allow ruling out confounding by tobacco smoking as a possible explanation for the observed associations.

Another potential explanation for the observed association of vitamin B6 with risk is reverse causation. In analysis stratified on smoking status, we observed a gradually stronger inverse association between vitamin B6 and risk of lung cancer in both former and current smokers when measured in blood drawn closer to diagnosis (Supplementary Figure 1, 3 and 4), results that would be consistent with pre-clinical metabolic changes due to the underlying disease progression. In contrast, no such clear relation by time from blood draw to diagnosis was observed for folate.

On the whole, we interpret these results as being consistent with a small benefit in terms of lung cancer risk for smokers who avoid low circulating concentrations of vitamin B6 and folate, but cannot rule out the possibility that residual confounding by tobacco exposure and/or reverse causation due to the disease progression, underlie their associations with risk.

The most notable difference between our results and those from the previous EPIC study was for never smokers, where vitamin B6 was positively associated with risk among for women in the current study, the opposite of what was observed in the previous EPIC study. The reason for this stark discrepancy is unclear and not explained by differences in study characteristics.

### **Study strengths and limitations**

Strengths of this study included the large study sample of 5,364 case-control pairs, the prospective study design and the use of pre-diagnostic plasma or serum, the centralized biochemical analysis with a robust quality control protocol, and the wide participation of cohorts from different geographical regions. We benefitted from over-sampling never and former smokers, as well as matching by history of tobacco exposure, thus allowing for well-powered stratified analysis. However, this oversampling also meant that the individual cohorts provided different proportions of subjects by smoking status. For instance, the WHI only contributed never smoking women and the ATBC study only contributed current smoking men. Another particular feature of this study was that circulating vitamin B6 and folate varied substantially across cohorts and continents. Indeed, US participants had about 65% higher B6 and almost three-fold the median folate concentrations of European and

Asian study participants (see Table 1). A likely explanation of this is that up to 50% of U.S. citizens aged 50 years or more regularly consume dietary supplements such as multi-vitamin pills,<sup>254</sup> and because of the folate fortification in the US.<sup>255</sup> Differences in baseline characteristics only partially accounted for these differences, and this means that we could not efficiently compare participants at extremes of the distribution of folate concentrations because the controls were matched to cases within the same cohort, and we were reluctant to break that matching.

### **Study implications and conclusions**

This study highlights the importance of re-evaluating promising associations of putative risk factors indicated in initial studies across multiple study populations in a coordinated fashion. The LC3 clearly demonstrates that it is feasible to gather cohorts from around the world and provide robust information on disease risk association before resorting to intervention studies that are costly and risk harming the study participants. Whilst our results were consistent with a modest decrease in lung cancer risk among former and current smokers who avoid low circulating concentrations of vitamin B6 and folate, we could not rule out the possibility that confounding due to tobacco or reverse causation may explain these associations. Given that any potential beneficial effect of vitamin B6 or folate—if real—is likely to be small, our findings do not support the conduct of additional prevention studies with the view of using B-vitamins for primary prevention of lung cancer.

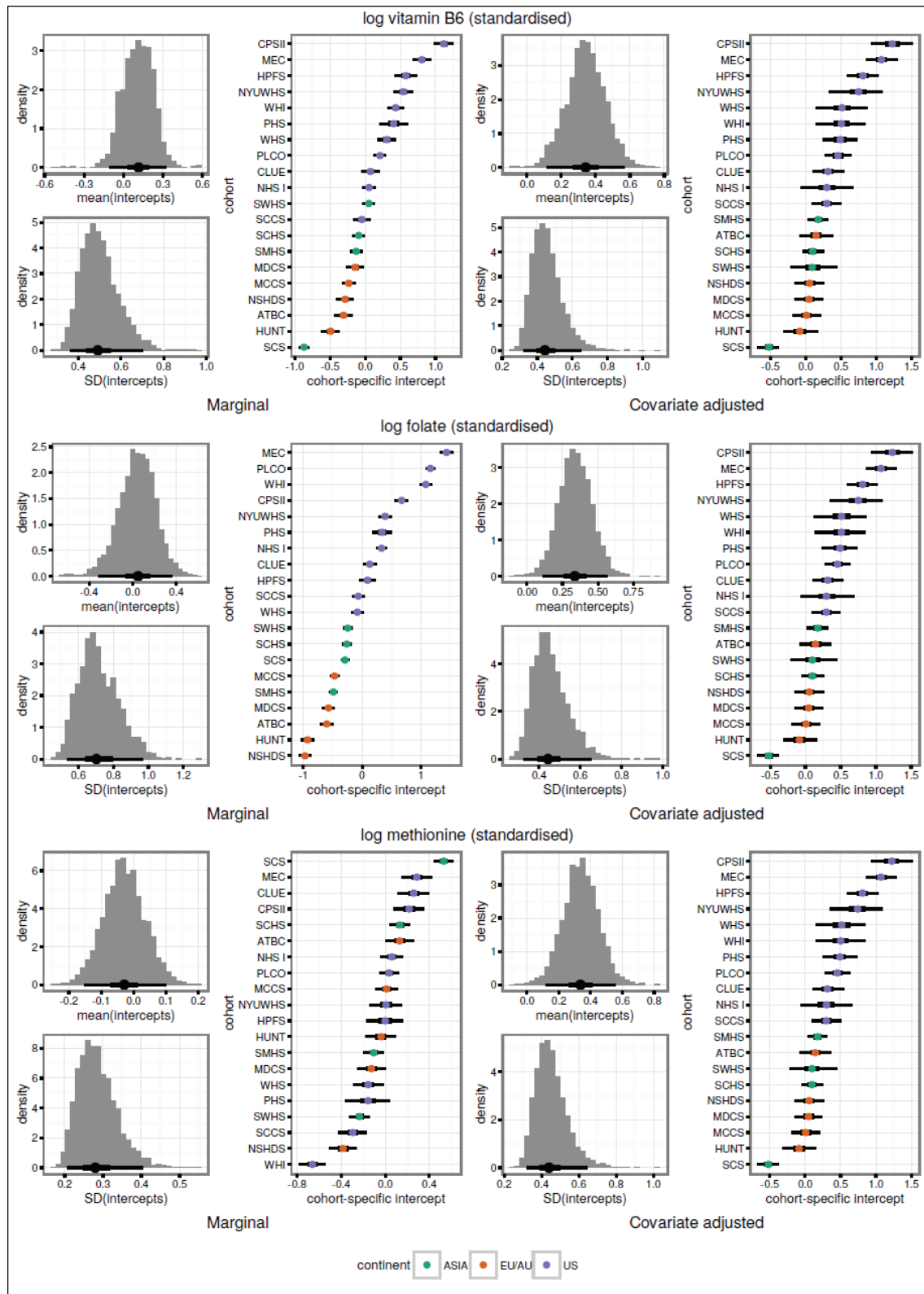
### **Funding**

The Lung Cancer Cohort Consortium (LC3) was supported by Grant NIH / NCI (n° 1U01CA155340-01). The funding organization had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript.

Drs Ueland and Middtun reports that they are members of the steering board of the nonprofit Foundation to Promote Research Into Functional Vitamin B12 Deficiency. No other disclosures were reported.

PB and MJ initiated, acquired the main funding, and designed this investigation. PMU and ØM led the laboratory analysis. AF and DM conducted the statistical analysis under supervision of MJ and PB. AF, MJ, JMY, VLS, SJW, DA, DM, RP, and PB drafted the first version of the manuscript. All authors were involved with collection of data, data interpretation, critical revisions of the paper, and approval of the final version. PB had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. PB and MJ are the overall coordinators of the LC3 consortium that was implemented in collaboration with the main investigators in the collaborating cohorts.

We thank the study participants and LC3 consortium members for their contributions in making this study possible.



**Figure 1.** Estimated between-cohort variation in mean concentration of vitamin B6, folate, and methionine. Before (left) and after (right) adjustment for age, sex, smoking status, and BMI. Histograms show the posterior distribution of the overall mean and standard deviation (SD) of the cohort-specific intercepts. The "caterpillar plots" show the estimated cohort-specific mean concentrations, along with 50% and 90% credible intervals (thick and thin bars, respectively).



**Table 1.** Baseline and clinical characteristics of study participants overall and per continent

Discrete variables	US cohorts			EU/AU cohorts			Asian cohorts			Overall		
	No.(%) of participants in group			No.(%) of participants in group			No.(%) of participants in group			No.(%) of participants in group		
	Cases (n=2400)	Matched controls (n=2400)		Cases (n=1189)	Matched controls (n=1189)		Cases (n=1775)	Matched controls (n=1775)		Cases (n=5364)	Matched controls (n=5364)	
<b>Sex</b>												
Men	991 (41.3%)	991 (41.3%)	688 (57.9%)	688 (57.9%)	688 (57.9%)	1229 (69.2%)	1229 (69.2%)	1229 (69.2%)	2908 (54.2%)	2908 (54.2%)	2908 (54.2%)	2908 (54.2%)
Women	1409 (58.7%)	1409 (58.7%)	501 (42.1%)	501 (42.1%)	501 (42.1%)	546 (30.8%)	546 (30.8%)	546 (30.8%)	2456 (45.8%)	2456 (45.8%)	2456 (45.8%)	2456 (45.8%)
<b>Smoking status</b>												
Never	569 (23.7%)	569 (23.7%)	156 (13.1%)	156 (13.1%)	156 (13.1%)	602 (33.9%)	602 (33.9%)	602 (33.9%)	1327 (24.7%)	1327 (24.7%)	1327 (24.7%)	1327 (24.7%)
Former	1007 (42%)	1007 (42%)	335 (28.2%)	335 (28.2%)	335 (28.2%)	176 (9.9%)	176 (9.9%)	176 (9.9%)	1518 (28.3%)	1518 (28.3%)	1518 (28.3%)	1518 (28.3%)
Current	824 (34.3%)	824 (34.3%)	698 (58.7%)	698 (58.7%)	698 (58.7%)	997 (56.2%)	997 (56.2%)	997 (56.2%)	2519 (47%)	2519 (47%)	2519 (47%)	2519 (47%)
<b>Education</b>												
less than high school	237 (9.9%)	215 (9%)	661 (55.6%)	661 (55.6%)	596 (50.2%)	898 (50.6%)	898 (49.7%)	898 (49.7%)	1797 (33.5%)	1696 (31.6%)	1696 (31.6%)	1696 (31.6%)
completed high school	357 (14.9%)	374 (15.6%)	159 (13.4%)	159 (13.4%)	180 (15.2%)	243 (13.7%)	230 (13%)	230 (13%)	759 (14.1%)	784 (14.6%)	784 (14.6%)	784 (14.6%)
vocational school	422 (17.6%)	435 (18.1%)	180 (15.2%)	180 (15.2%)	200 (16.8%)	289 (16.3%)	279 (15.7%)	279 (15.7%)	891 (16.6%)	914 (17%)	914 (17%)	914 (17%)
some college	402 (16.8%)	393 (16.4%)	107 (9%)	107 (9%)	129 (10.9%)	171 (9.6%)	196 (11%)	196 (11%)	680 (12.7%)	718 (13.4%)	718 (13.4%)	718 (13.4%)
college graduate	357 (14.9%)	319 (13.3%)	63 (5.3%)	63 (5.3%)	64 (5.4%)	104 (5.9%)	113 (6.4%)	113 (6.4%)	524 (9.8%)	496 (9.2%)	496 (9.2%)	496 (9.2%)
graduate studies	574 (23.9%)	637 (26.5%)	10 (0.8%)	10 (0.8%)	8 (0.7%)	62 (3.5%)	65 (3.7%)	65 (3.7%)	646 (12%)	710 (13.2%)	710 (13.2%)	710 (13.2%)
Unknown	51 (2.1%)	27 (1.1%)	8 (0.7%)	8 (0.7%)	10 (0.8%)	8 (0.5%)	9 (0.5%)	9 (0.5%)	67 (1.2%)	46 (0.9%)	46 (0.9%)	46 (0.9%)
<b>Alcohol use (any type)</b>												
Never	250 (10.4%)	252 (10.5%)	292 (24.6%)	292 (24.6%)	286 (24.1%)	946 (53.3%)	974 (54.9%)	974 (54.9%)	1488 (27.7%)	1512 (28.2%)	1512 (28.2%)	1512 (28.2%)
Ever	2013 (83.9%)	2041 (85%)	643 (54.1%)	643 (54.1%)	643 (54.1%)	829 (46.7%)	801 (45.1%)	801 (45.1%)	3485 (65%)	3485 (65%)	3485 (65%)	3485 (65%)
Unknown	137 (5.7%)	107 (4.5%)	254 (21.4%)	254 (21.4%)	260 (21.9%)	0 (0%)	0 (0%)	0 (0%)	391 (7.3%)	367 (6.8%)	367 (6.8%)	367 (6.8%)
<b>Body Mass index<sup>a</sup></b>												
<18.5	30 (1.3%)	31 (1.3%)	14 (1.2%)	14 (1.2%)	9 (0.8%)	157 (8.8%)	113 (6.4%)	113 (6.4%)	201 (3.7%)	153 (2.9%)	153 (2.9%)	153 (2.9%)
18.5-25	1088 (45.3%)	1020 (42.5%)	521 (43.8%)	521 (43.8%)	435 (36.6%)	1203 (67.8%)	1192 (67.2%)	1192 (67.2%)	2812 (52.4%)	2647 (49.3%)	2647 (49.3%)	2647 (49.3%)
25-30	841 (35%)	858 (35.8%)	468 (39.4%)	468 (39.4%)	536 (45.1%)	369 (20.8%)	424 (23.9%)	424 (23.9%)	1678 (31.3%)	1818 (33.9%)	1818 (33.9%)	1818 (33.9%)
≥30	378 (15.8%)	430 (17.9%)	185 (15.6%)	185 (15.6%)	207 (17.4%)	46 (2.6%)	46 (2.6%)	46 (2.6%)	609 (11.4%)	683 (12.7%)	683 (12.7%)	683 (12.7%)
Unknown	63 (2.6%)	61 (2.5%)	1 (0.1%)	1 (0.1%)	2 (0.2%)	0 (0%)	0 (0%)	0 (0%)	64 (1.2%)	63 (1.2%)	63 (1.2%)	63 (1.2%)
<b>Continuous variables, median (5th-95th percentile)</b>												
Age at recruitment (years)	60 (42-74)	60 (42-74)	60 (45-70)	60 (45-70)	60 (45-70)	60 (46-72)	60 (46-72)	60 (46-72)	60 (44-72)	60 (44-72)	60 (44-72)	60 (44-72)
Blood concentrations for biomarkers												
Vitamin B6 (Pyridoxal 5'-phosphate), nmol/L	47.6 (15.2-266)	49.9 (16.4-271)	28.9 (12.6-127)	28.9 (12.6-127)	30.9 (13.3-102)	29.0 (11.1-114)	31.2 (12.4-119)	31.2 (12.4-119)	35.2 (12.6-205)	37.1 (13.9-197)	37.1 (13.9-197)	37.1 (13.9-197)
Vitamin B9 (folate), nmol/L	32.9 (8.3-114.2)	33.7 (8.1-114.9)	10.2 (4.4-30.1)	10.2 (4.4-30.1)	10.8 (4.9-32.9)	13.9 (6.2-40.5)	15.1 (6.7-40.7)	15.1 (6.7-40.7)	17.0 (6.0-91.8)	17.9 (6.3-90.3)	17.9 (6.3-90.3)	17.9 (6.3-90.3)
Methionine, μmol/L	25.8 (18.4-39.6)	26.2 (18.8-39.7)	26.1 (19.1-37.8)	26.1 (19.1-37.8)	26.3 (18.7-37.8)	27.3 (18.0-42.4)	27.4 (18.2-41.9)	27.4 (18.2-41.9)	26.3 (18.4-40.4)	26.6 (18.6-40.3)	26.6 (18.6-40.3)	26.6 (18.6-40.3)
<b>Clinical characteristics, case participants only</b>												
Age at diagnosis, median (range), years	70 (55-83)		69 (54-79)	69 (54-79)		69 (52-80)			69.8 (53.6-82.0)			

Time from blood draw to diagnosis (years)	5.2 (1-15.5)	10.0 (1.5-16.0)	5.8 (0.7-16.5)	6.3 (1.0-16.0)
Hystology, No. (%)				
Large cell carcinoma	112 (4.6%)	46 (4%)	16 (1%)	174 (3.3%)
Small cell carcinoma	245 (10.4%)	150 (12.5%)	99 (5.5%)	492 (9.2%)
Squamous cell carcinoma	291 (11.9%)	231 (19.5%)	319 (17.9%)	836 (15.5%)
Adenocarcinoma	1034 (42.7%)	419 (34.5%)	615 (34.6%)	2056 (38.4%)
Missing / Unknown	735 (31.4%)	357 (29.5%)	726 (41%)	1806 (33.6%)

<sup>a</sup> Body mass index is calculated as weight in kilograms divided by height in meters squared

**Table 2.** Odds ratios of lung cancer for circulating concentrations of vitamin B6

		Odds Ratio [95% Confidence Interval]								
LC3 participants	Quartile (Range)	Case/Controls participants	Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L] <sup>d</sup>	Cases compared to matched controls, unadjusted <sup>a</sup>	Never Smokers <sup>c</sup>	Former Smokers <sup>c</sup>	Current smokers <sup>b</sup>	Men <sup>b</sup>	Women <sup>b</sup>	
				(n=5364/5364) <sup>a</sup>	(n=1327/1327) <sup>c</sup>	(n = 1518/1518) <sup>c</sup>	(n = 2519/2519) <sup>b</sup>	(n=2908/2908) <sup>b</sup>	(n=2456/2456) <sup>b</sup>	
<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>										
United States (n=11 cohorts)	1 (5.08-29.89)	629/600	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (29.90-49.90)	618/600	0.98 (0.83 - 1.15)	1.00 (0.84 - 1.18)	1.45 (0.99 - 2.13)	0.88 (0.68 - 1.14)	0.95 (0.73 - 1.24)	0.82 (0.63 - 1.06)	1.14 (0.91 - 1.41)	
	3 (49.91-90.05)	574/600	0.90 (0.77 - 1.07)	0.93 (0.78 - 1.10)	1.19 (0.82 - 1.72)	0.87 (0.67 - 1.13)	0.89 (0.67 - 1.17)	0.77 (0.59 - 1.00)	1.04 (0.84 - 1.30)	
	4 > 90.05	579/600	0.91 (0.76 - 1.07)	0.95 (0.8 - 1.13)	1.19 (0.82 - 1.72)	0.90 (0.69 - 1.17)	0.85 (0.63 - 1.15)	0.75 (0.56 - 0.99)	1.11 (0.89 - 1.38)	
<i>P for trend<sup>h</sup></i>										
0.14										
<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>e</sup></b>										
Europe /Australia (n=5 cohorts)	1 (4.37-22.41)	378/297	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.42-30.90)	269/297	0.70 (0.56 - 0.88)	0.73 (0.58 - 0.92)	1.32 (0.59 - 2.98)	0.97 (0.59 - 1.60)	0.62 (0.47 - 0.82)	0.73 (0.54 - 0.99)	0.71 (0.49 - 1.03)	
	3 (30.91-44.95)	258/298	0.65 (0.51 - 0.83)	0.69 (0.54 - 0.88)	1.19 (0.56 - 2.54)	0.61 (0.37 - 0.99)	0.71 (0.52 - 0.97)	0.74 (0.53 - 1.02)	0.61 (0.42 - 0.89)	
	4 > 44.96	284/297	0.73 (0.57 - 0.92)	0.78 (0.62 - 1.00)	1.57 (0.77 - 3.19)	0.73 (0.46 - 1.17)	0.76 (0.55 - 1.04)	0.68 (0.50 - 0.94)	0.91 (0.64 - 1.31)	
<i>P for trend<sup>h</sup></i>										
0.12										
<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>f</sup></b>										
Asia (n=4 cohorts)	1 (4.81-20.49)	528/443	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (20.50-30.82)	440/444	0.78 (0.64 - 0.96)	0.90 (0.73 - 1.10)	1.33 (0.80 - 2.21)	0.67 (0.32 - 1.41)	0.92 (0.72 - 1.17)	0.89 (0.71 - 1.11)	1.29 (0.74 - 2.25)	
	3 (31.23-46.99)	398/444	0.67 (0.55 - 0.84)	0.79 (0.63 - 0.98)	1.59 (0.95 - 2.68)	0.60 (0.31 - 1.17)	0.66 (0.50 - 0.86)	0.68 (0.52 - 0.87)	1.45 (0.84 - 2.51)	
	4 > 47.00	409/444	0.68 (0.55 - 0.87)	0.82 (0.65 - 1.02)	1.55 (0.99 - 2.72)	0.54 (0.28 - 1.02)	0.66 (0.48 - 0.93)	0.75 (0.58 - 0.98)	1.34 (0.79 - 2.29)	
<i>P for trend<sup>h</sup></i>										
0.002										
<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>g</sup></b>										
All cohorts (n=20 cohorts)	1 (4.37-23.93)	1501/1341	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (23.94-37.05)	1319/1341	0.85 (0.76 - 0.95)	0.90 (0.81 - 1.01)	1.53 (1.15 - 2.05)	0.73 (0.58 - 0.93)	0.86 (0.74 - 1.00)	0.77 (0.66 - 0.89)	1.14 (0.96 - 1.37)	
	3 (37.06-62.50)	1254/1341	0.80 (0.72 - 0.90)	0.87 (0.77 - 0.97)	1.52 (1.15 - 2.00)	0.66 (0.53 - 0.84)	0.83 (0.70 - 0.98)	0.77 (0.66 - 0.91)	1.03 (0.86 - 1.23)	
	4 > 62.51	1290/1341	0.81 (0.72 - 0.92)	0.88 (0.78 - 1.00)	1.51 (1.14 - 2.01)	0.72 (0.57 - 0.91)	0.79 (0.65 - 0.95)	0.74 (0.62 - 0.89)	1.10 (0.92 - 1.31)	
<i>P for trend<sup>h</sup></i>										
9x10 <sup>-4</sup>										

a) Assessed by analysing lung cancer cases and their individually matched controls by conditional logistic regression, conditioning on individual case set.

b) Assessed by analysing lung cancer cases and their individually matched controls by conditional logistic regression, conditioning on individual case set, and adjusting for circulating cotinine (in quartiles) and education (in 7 categories).

c) Assessed by analysing lung cancer cases and their individually matched controls by conditional logistic regression, conditioning on individual case set, and adjusting for education (in 7 categories).

d) Quartile cut-off points were determined based on the blood concentration distribution of each biomarker for 2400 individually matched controls (US cohorts; n=11).

- e) Quartile cut-off points were determined based on the blood concentration distribution of each biomarker for 1189 individually matched controls (European/Australian cohorts; n=5).
- f) Quartile cut-off points were determined based on the blood concentration distribution of each biomarker for 1775 individually matched controls (Asian cohorts; n=4).
- g) Quartile cut-off points were determined based on the blood concentration distribution of each biomarker for 5364 individually matched controls (pooled cohorts; n=20).
- h) P for trend assessed by the base 2 logarithm of the circulating concentrations.

**Table 3.** Odds ratios of lung cancer for circulating concentrations of folate

LC3 participants	Quartile (Range)	Case/Controls participants	Odds Ratio (95% Confidence Interval)								
			Cases compared to matched controls, unadjusted <sup>a</sup> (n=5364/5364) <sup>a</sup>	Cases compared to matched controls, adjusted for cotinine and education <sup>b</sup> (n=5364/5364) <sup>b</sup>	Never Smokers <sup>c</sup> (n=1327/1327) <sup>c</sup>	Former Smokers <sup>c</sup> (n = 1518/1518) <sup>c</sup>	Current smokers <sup>b</sup> (n = 2519/2519) <sup>b</sup>	Men <sup>b</sup> (n=2908/2908) <sup>b</sup>	Women <sup>b</sup> (n=2456/2456) <sup>b</sup>		
United States (n=11 cohorts)	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>										
	1 (1.70-17.45)	633/600	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	
	2 (17.46-33.35)	578/600	0.90 (0.77 - 1.07)	0.93 (0.78 - 1.10)	0.83 (0.54 - 1.29)	0.95 (0.72 - 1.24)	0.92 (0.72 - 1.18)	0.86 (0.66 - 1.12)	0.98 (0.79 - 1.22)		
	3 (33.36-59.10)	598/600	0.93 (0.78 - 1.10)	0.95 (0.80 - 1.14)	0.87 (0.56 - 1.36)	0.86 (0.66 - 1.13)	1.05 (0.78 - 1.42)	0.92 (0.69 - 1.22)	0.98 (0.78 - 1.24)		
	4 > 59.10	591/600	0.91 (0.75 - 1.10)	0.95 (0.78 - 1.15)	0.92 (0.57 - 1.47)	0.87 (0.65 - 1.16)	1.02 (0.72 - 1.45)	0.78 (0.56 - 1.07)	1.07 (0.84 - 1.37)		
	<i>P for trend<sup>h</sup></i>	0.27	0.52	0.84	0.27	0.88	0.44	0.85			
Europe /Australia (n=5 cohorts)	<b>Folate (Vitamin B9) [nmol/L]<sup>e</sup></b>										
	1 (0.24-7.78)	348/297	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)		
	2 (7.79-10.77)	291/297	0.82 (0.65 - 1.03)	0.84 (0.66 - 1.06)	1.30 (0.65 - 2.60)	0.71 (0.44 - 1.14)	0.81 (0.60 - 1.10)	0.82 (0.59 - 1.12)	0.85 (0.59 - 1.22)		
	3 (10.78-15.98)	282/298	0.78 (0.61 - 0.98)	0.80 (0.63 - 1.02)	0.65 (0.33 - 1.25)	0.69 (0.44 - 1.10)	0.90 (0.65 - 1.23)	0.76 (0.55 - 1.04)	0.85 (0.58 - 1.23)		
	4 > 15.98	268/297	0.73 (0.57 - 0.93)	0.77 (0.60 - 0.99)	0.98 (0.49 - 1.97)	0.58 (0.36 - 0.93)	0.83 (0.60 - 1.16)	0.69 (0.49 - 0.98)	0.87 (0.60 - 1.27)		
	<i>P for trend<sup>h</sup></i>	0.007	0.02	0.89	0.007	0.37	0.02	0.47			
Asia (n=4 cohorts)	<b>Folate (Vitamin B9) [nmol/L]<sup>f</sup></b>										
	1 (0.17-10.59)	525/443	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)		
	2 (10.60-15.00)	461/444	0.84 (0.70 - 1.02)	0.88 (0.72 - 1.06)	0.96 (0.65 - 1.40)	0.40 (0.21 - 0.74)	0.96 (0.75 - 1.23)	0.85 (0.68 - 1.06)	1.01 (0.68 - 1.5)		
	3 (15.01-21.00)	366/444	0.67 (0.55 - 0.81)	0.73 (0.59 - 0.89)	0.76 (0.52 - 1.10)	0.45 (0.23 - 0.85)	0.80 (0.65 - 1.12)	0.75 (0.59 - 0.95)	0.78 (0.53 - 1.14)		
	4 >21.01	423/444	0.77 (0.63 - 0.94)	0.84 (0.68 - 1.03)	0.89 (0.61 - 1.29)	0.65 (0.35 - 1.18)	0.84 (0.64 - 1.11)	0.74 (0.58 - 0.94)	1.12 (0.76 - 1.65)		
	<i>P for trend<sup>h</sup></i>	0.001	0.02	0.26	0.26	0.09	0.008	0.97			
All cohorts (n=20 cohorts)	<b>Folate (Vitamin B9) [nmol/L]<sup>g</sup></b>										
	1 (0.17-10.92)	1489/1341	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)		
	2 (10.93-17.89)	1331/1341	0.87 (0.78 - 0.97)	0.90 (0.81 - 1.01)	0.87 (0.67 - 1.13)	0.74 (0.58 - 0.95)	0.96 (0.83 - 1.12)	0.90 (0.77 - 1.04)	0.92 (0.77 - 1.10)		
	3 (17.90-34.92)	1243/1341	0.79 (0.70 - 0.89)	0.84 (0.74 - 0.95)	0.81 (0.62 - 1.06)	0.73 (0.57 - 0.93)	0.88 (0.74 - 1.04)	0.77 (0.65 - 0.91)	0.94 (0.78 - 1.13)		
	4 > 34.97	1301/1341	0.80 (0.70 - 0.92)	0.86 (0.74 - 0.99)	0.86 (0.63 - 1.17)	0.66 (0.51 - 0.85)	0.97 (0.77 - 1.21)	0.75 (0.61 - 0.93)	0.97 (0.79 - 1.19)		
	<i>P for trend<sup>h</sup></i>	2x10 <sup>-4</sup>	0.01	0.37	0.01	0.15	0.001	0.66			

a) Assessed by analysing lung cancer cases and their individually matched controls by conditional logistic regression, conditioning on individual case set.

b) Assessed by analysing lung cancer cases and their individually matched controls by conditional logistic regression, conditioning on individual case set, and adjusting for circulating cotinine (in quartiles) and education (in 7 categories).

c) Assessed by analysing lung cancer cases and their individually matched controls by conditional logistic regression, conditioning on individual case set, and adjusting for education (in 7 categories).

d) Quartile cut-off points were determined based on the blood concentration distribution of each biomarker for 2400 individually matched controls (US cohorts; n=11).

e) Quartile cut-off points were determined based on the blood concentration distribution of each biomarker for 1189 individually matched controls (European/Australian cohorts; n=5).

f) Quartile cut-off points were determined based on the blood concentration distribution of each biomarker for 1775 individually matched controls (Asian cohorts; n=4).

g) Quartile cut-off points were determined based on the blood concentration distribution of each biomarker for 5364 individually matched controls (pooled cohorts; n=20).

h) P for trend assessed by the base 2 logarithm of the circulating concentrations.

**Table 4.** Odds ratios of lung cancer for circulating concentrations of methionine

## Odds Ratio (95% Confidence Interval)

LC3 participants	Quartile (Range)	Case/Controls participants	Cases compared to matched controls, adjusted for cotinine and education <sup>b</sup>		Never Smokers <sup>c</sup> (n=1327/1327) <sup>c</sup>	Former Smokers <sup>c</sup> (n = 1518/1518) <sup>c</sup>	Current smokers <sup>b</sup> (n = 2519/2519) <sup>b</sup>	Men <sup>b</sup> (n=2908/2908) <sup>b</sup>	Women <sup>b</sup> (n=2456/2456) <sup>b</sup>
			Cases compared to matched controls, unadjusted <sup>a</sup>	Cases compared to matched controls, adjusted for cotinine and education <sup>b</sup>					
United States (n=11 cohorts)	Methionine [ $\mu\text{mol/L}$ ] <sup>d</sup>								
	1 (11.83-22.77)	632/600	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.78-26.16)	617/600	0.97 (0.82 - 1.14)	0.97 (0.82 - 1.14)	1.10 (0.80 - 1.51)	0.82 (0.62 - 1.08)	1.02 (0.78 - 1.35)	0.90 (0.68 - 1.19)	1.00 (0.82 - 1.23)
	3 (26.17-30.47)	591/600	0.92 (0.78 - 1.09)	0.93 (0.78 - 1.10)	1.04 (0.73 - 1.48)	0.78 (0.60 - 1.01)	1.01 (0.75 - 1.36)	0.88 (0.67 - 1.16)	0.95 (0.77 - 1.19)
	4 > 30.47	560/600	0.87 (0.73 - 1.03)	0.87 (0.73 - 1.04)	1.05 (0.68 - 1.62)	0.79 (0.60 - 1.02)	0.84 (0.63 - 1.13)	0.68 (0.52 - 0.89)	1.10 (0.86 - 1.39)
<i>P for trend</i> <sup>h</sup>									
Europe /Australia (n=5 cohorts)	Methionine [ $\mu\text{mol/L}$ ] <sup>e</sup>								
	1 (10.63-22.81)	295/297	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.82-26.29)	316/297	1.08 (0.86 - 1.35)	1.07 (0.85 - 1.35)	1.01 (0.54 - 1.90)	0.88 (0.57 - 1.35)	1.21 (0.88 - 1.65)	1.08 (0.76 - 1.52)	1.07 (0.78 - 1.48)
	3 (26.30-29.91)	282/298	0.94 (0.74 - 1.20)	0.91 (0.71 - 1.16)	0.82 (0.43 - 1.57)	1.06 (0.67 - 1.65)	0.87 (0.62 - 1.20)	0.94 (0.67 - 1.33)	0.87 (0.60 - 1.24)
	4 > 29.91	296/297	0.99 (0.78 - 1.26)	1.00 (0.78 - 1.27)	0.96 (0.49 - 1.89)	1.11 (0.72 - 1.70)	0.95 (0.68 - 1.32)	1.02 (0.74 - 1.42)	0.96 (0.65 - 1.41)
<i>P for trend</i> <sup>h</sup>									
Asia (n=4 cohorts)	Methionine [ $\mu\text{mol/L}$ ] <sup>f</sup>								
	1 (12.32-23.17)	451/443	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (23.18-27.39)	440/444	0.99 (0.81 - 1.20)	1.02 (0.84 - 1.24)	0.85 (0.63 - 1.16)	0.70 (0.38 - 1.29)	1.32 (0.98 - 1.77)	1.15 (0.89 - 1.49)	0.86 (0.62 - 1.17)
	3 (27.39-32.22)	420/444	0.92 (0.75 - 1.12)	0.93 (0.76 - 1.15)	0.99 (0.71 - 1.38)	0.73 (0.40 - 1.32)	1.00 (0.74 - 1.34)	0.94 (0.73 - 1.22)	1.02 (0.71 - 1.47)
	4 > 32.22	464/444	1.02 (0.82 - 1.22)	1.04 (0.85 - 1.26)	1.11 (0.79 - 1.57)	0.59 (0.32 - 1.10)	1.16 (0.88 - 1.53)	1.13 (0.88 - 1.45)	0.86 (0.6 - 1.22)
<i>P for trend</i> <sup>h</sup>									
All cohorts (n=20 cohorts)	Methionine [ $\mu\text{mol/L}$ ] <sup>g</sup>								
	1 (10.63-22.89)	1377/1341	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.90-27.40)	1382/1341	1.00 (0.90 - 1.12)	1.01 (0.91 - 1.13)	0.99 (0.80 - 1.21)	0.88 (0.72 - 1.09)	1.12 (0.95 - 1.32)	1.02 (0.87 - 1.21)	1.00 (0.86 - 1.16)
	3 (27.40-31.00)	1303/1341	0.93 (0.83 - 1.05)	0.94 (0.84 - 1.05)	0.98 (0.78 - 1.22)	0.80 (0.65 - 0.99)	1.01 (0.84 - 1.20)	0.89 (0.76 - 1.05)	1.00 (0.85 - 1.17)
	4 > 31.01	1302/1341	0.93 (0.83 - 1.04)	0.95 (0.85 - 1.07)	1.04 (0.81 - 1.33)	0.83 (0.67 - 1.03)	0.99 (0.83 - 1.17)	0.92 (0.79 - 1.08)	0.99 (0.83 - 1.18)
<i>P for trend</i> <sup>h</sup>									

a) Assessed by analysing lung cancer cases and their individually matched controls by conditional logistic regression, conditioning on individual case set.

b) Assessed by analysing lung cancer cases and their individually matched controls by conditional logistic regression, conditioning on individual case set, and adjusting for circulating cotinine (in quartiles) and education (in 7 categories).

c) Assessed by analysing lung cancer cases and their individually matched controls by conditional logistic regression, conditioning on individual case set, and adjusting for education (in 7 categories).

d) Quartile cut-off points were determined based on the blood concentration distribution of each biomarker for 2400 individually matched controls (US cohorts; n=11).

e) Quartile cut-off points were determined based on the blood concentration distribution of each biomarker for 1189 individually matched controls (Asian cohorts; n=4).

f) Quartile cut-off points were determined based on the blood concentration distribution of each biomarker for 1775 individually matched controls (European/Australian cohorts; n=5).

g) Quartile cut-off points were determined based on the blood concentration distribution of each biomarker for 5364 individually matched controls (pooled cohorts; n=20).

h) *P for trend* assessed by the base 2 logarithm of the circulating concentrations.

# **CHAPTER VI:**

## ***DISCUSSION***

## **The one-carbon metabolism and smoking-related cancers**

### **Recall on objectives**

Over the past two decades, the role of one-carbon metabolism (OCM) in cancer has received considerable attention. Circulating B-vitamins and related analytes have been assessed prospectively for multiple cancer sites including colorectal<sup>256-259</sup>, prostate<sup>237 238</sup>, and breast cancer<sup>260-263</sup>, although the evidence is inconsistent. To date, data examining associations of OCM deficiency/ imbalance and smoking related cancers are sparse, and my thesis aims to fill this knowledge gap. In order to achieve this, circulating OCM biomarkers were examined in relation to three specific cancer sites that differed in their strength of association with smoking.

### **Summary of the results**

#### *One-carbon metabolism and head and neck cancer etiology and outcome (Chapter III)*

To the best of our knowledge, this was the first prospective study assessing the association between circulating OCM biomarkers and risk of the head and neck cancer (HNC). Only a few retrospective case-control studies had been published and reported lower circulating concentrations of folate and higher concentrations of homocysteine in cases than in controls<sup>203-206</sup>. In this study, we observed that subjects in the top quartile of circulating concentrations of homocysteine had a two-fold increase in risk of HNC compared with those in the bottom quartile. The positive association of homocysteine was present both among never-smokers and non-alcohol consumers.

Furthermore, we observed that participants with high circulating concentrations of folate had overall lower risk of HNC. Because folate and homocysteine are inversely correlated in the OCM pathway<sup>135</sup>, we cannot exclude the possibility that the folate association with risk is underlying the positive association of homocysteine with HNC risk. Mutually adjusted conditional analysis (data not shown) indicated that these risk associations were not independent, homocysteine accounting for the association of folate, but not vice-versa. However, the inverse folate association with HNC risk was not as consistent as that of homocysteine, and when performing different sensitivity analyses, the positive homocysteine association with risk was not materially affected whereas the folate



association was attenuated and no longer significant. This finding might not support for a decrease of HNC risk in participants having elevated dietary intake of folate, as seen in previous epidemiological studies<sup>69 200-202</sup>.

Overall, the results from Chapter II indicate that circulating homocysteine affects the risk of developing squamous cell carcinoma of the head and neck.

#### *One-carbon metabolism and renal cell carcinoma etiology and outcome (Chapter IV)*

Evidence on the role of vitamin B6 in RCC is limited, , and to date, only one prospective study assessing circulating concentrations of vitamin B6 and RCC risk has been published<sup>212</sup>, and did not report any clear association.

In contrast, our study showed that subjects from the EPIC cohort with elevated vitamin B6 concentrations had lower risk of RCC, in a dose-response fashion, resulting in two to three-fold risk differences between the top and the bottom quartiles of the population. We also observed a clear improved survival among cases with higher vitamin B6 concentrations. Subsequent replication of the EPIC results using the independent MCCS cohort from Melbourne, Australia, yielded nearly identical associations for both RCC risk as well as for cancer mortality.

Because the RCC study from the EPIC cohort and the previous study from the ATBC study are different in term of study design (population-based observational study vs trial) and populations sources (10 European countries vs Finish men smokers) and also because these two studies used different biological measurement methods to measure circulating vitamin B6 (liquid chromatography/tandem mass spectrometry vs the tyrosine decarboxylase assay), we cannot exclude that these study differences might explain this discordance in the findings.

One of the main questions that arose following the findings from the RCC study was whether the association of circulating vitamin B6 with survival was also present among participants having blood measurement at diagnosis. In order to investigate this, our group analyzed plasma from 630 newly diagnosed RCC cases from Central and Eastern Europe (Muller DC *et al*, in press Plos One, 09/2015) and found that the hazard of death was three times lower among those in the highest compared with the lowest fourth of vitamin B6 concentrations after taking into account disease stage.

Together, these studies strongly suggest that circulating concentrations of vitamin B6 are inversely related to both RCC risk and mortality.

*One-carbon metabolism and lung cancer etiology in the LC3 (Chapter V)*

This large consortium of lung cancers, the LC3 study, was based on previous results from EPIC cohort where a strong dose-response association between circulating vitamin B6 and lung cancer risk was reported including among never smokers<sup>13</sup>.

In the current study, the LC3 study, we observed a weak inverse association, with no clear trend, between circulating concentrations of folate and B6 and lung cancer risk, and this association was mainly present among ever smoking study participants. Similarly, men and former smokers who were deficient in circulating vitamin B6 and folate showed a marginal increased risk of lung cancer compared to those with normal levels. The observed results from LC3 study did not confirm the reported results from the previous EPIC lung study. However, it is not straightforward to compare our results with those of the EPIC lung study because the EPIC cohort was conducted mainly in Western European countries whereas the LC3 study is more diverse, and included four continents, several ethnic backgrounds and sex-specific cohorts. It would seem clear that the discordant results between the studies cannot be readily explained by differences in baseline characteristics such as sex or smoking status. It is possible that unobserved differences between the cohorts might partially account for the discrepant results.

One particular feature of the LC3 study is that circulating vitamin B6 and folate varied substantially across cohorts and continents; American participants displayed almost three-fold higher median folate concentrations and 1.5-fold higher median vitamin B6 concentrations compare to both the European and Asian participants. This drastic variation complicated comparisons, as well as pooling of data across continents and cohorts. In addition, the sizable sample size of 5,364 case-control pairs quickly diminished when stratifying by continent and smoking status resulting in reduced statistical power in some important subgroups such as never smoking men.

Furthermore, when looking for deficiency, we noted a relatively small proportion among US participants being deficient in circulating folate (2% being below 7 nmol/L) and vitamin B6 (10% being below 20 nmol/L) compare to European/Australian and Asian participants. Speculatively, the lack of variability from deficiency to excess of these biomarkers might be

a potential explanation of the null association observed among US participants when evaluating the relation of deficiency with lung cancer risk, and therefore cannot rule out the possibility that low circulating of folate and vitamin B6 may lead to stronger increased risk of lung cancer compared to what we found in the LC3 study.

Results from the LC3 study suggest that ever smoking study participants with elevated circulating vitamin B6 and folate had a slight decreased risk in lung cancer, and they should avoid deficiency both in circulating folate and vitamin B6, but cannot rule out that tobacco exposure may explain this result.

## **Observations across the studies**

### *Tobacco exposure*

Because the control of confounding by smoking depends on accurate smoking history, and because some misclassification of the true smoking history is unavoidable, investigations of smoking related-cancers overall are difficult to conduct. Moreover, the proportion of cancers explained by smoking, i.e. the attributable fraction, differs among cancer histology and cancer sites, as for the three cancers examined in this thesis. While tobacco smoking dramatically increases the risk of lung cancer, the risk increase is moderate for HNC and weaker for RCC. As such, residual confounding by smoking is difficult to exclude, in particular for lung cancer.

In HNC study, the association between homocysteine and HNC risk was observed among both never and current smokers, suggesting that residual confounding due to tobacco smoking might not explain the association with risk.

In RCC study, the test of interaction between smoking exposures and vitamin B6 revealed that the inverse association between circulating concentrations of vitamin B6 and RCC risk was prominent among current compared to never- and former smokers, although a non-significant inverse association among the two latter strata was observed. However, current smokers have about 30 to 50% higher risk of RCC compared to never smokers<sup>1 7</sup>, and given that the association of circulating vitamin B6 indicated 2 to 3-fold increased risk among

current smokers compared to never smokers, it would seem unlikely that confounding by smoking can explain the observed results.

In LC3 study, associations were mainly observed among ever smokers. We attempted to control for confounding by tobacco exposure by matching by detailed smoking status. In addition, circulating cotinine, which is an indicator of current smoking intensity and is known to explain lung cancer risk beyond that captured by questionnaire information<sup>253</sup>, was also included to control tobacco smoking among current smokers in each study of this dissertation. Even though we took all possible measure to account for residual confounding by tobacco exposure, we cannot completely rule out the possibility that it may still partially account for the observed associations of lung cancer.

While we observed an inverse association between circulating vitamin B6 and lung cancer risk among ever smokers, the vitamin B6 association surprisingly tend to be positively associated to lung cancer among never smokers. We do not know exactly the reason for this discrepant results but this could be due to the difference of comparability between never smoking lung cancer cases and controls. While ever smoking cases and controls shared a similar strong risk factor that lead to lung cancer development, we might miss comparability of exposures between never smoking lung cancer cases and controls. Lung cancer among never smokers is low<sup>30</sup> and to date, the descriptive epidemiology of never smokers is limited, although several risk factors have been suggested including second hand smoke<sup>30 33</sup>, occupational exposure<sup>32 264</sup>, asbestos exposure<sup>265</sup>, and genetic factors<sup>266</sup>. It is possible that never smoker participants diagnosed with lung cancer in LC3 study were related to these previous or unknown risk factors, which may account for these surprising findings among non-smokers. However, we did not have such information in the LC3 study, and therefore this complicates the comparability between never smoking lung cancers cases and controls, even though they were matched by relevant confounding including age, sex and cohort. Assembling further studies on never smoking lung cancers will ensure a better assessment of their risk factors.

#### *Other potential confounders*

A potential explanation of the results observed throughout the studies of this thesis is that underlying preclinical disease may affect concentrations of OCM biomarkers. However, the

associations observed in HNC and RCC studies were stable over the 10-year of follow-up after blood collection. In the same two studies, our estimates did not change appreciably after excluding cases with their matched controls who were diagnosed within 1 or 2 years after blood draw, which would seem to exclude any possible reverse causation bias. Furthermore, the observed associations throughout the three studies of this thesis are unlikely to be explained by the other OCM biomarkers because our estimates were not affected after controlling for them, either one-at-time or simultaneously.

As mentioned in the beginning of this thesis, major sources of one-carbon nutrients and related vitamins are many and include fruits and green leafy vegetables (folate) as well as fortified cereals and whole grains (vitamin B6). Thus, one could ask whether food intake affects circulating OCM biomarkers or cancer development. As example from the RCC study, we observed no association between vitamin B6 estimated from food frequency questionnaire (FFQ) and risk of RCC, in contrast to the strong association from circulating concentrations. There are suggestive explanations that could explain this discrepancy. First, is that blood measurement reflects more accurately vitamin B6 intake than estimates from self-reported questionnaires. Although measurement correction error models have been developed<sup>267</sup>, misclassification of dietary intake is still a major concern in dietary assessment. Moreover, in RCC study, the correlation between vitamin B6 intake from FFQ and plasma vitamin B6 was very modest (0.17), an observation very similar to multiple previous studies<sup>268 269</sup>. Also, adjusting for fruit, vegetables, nuts and seed, that are rich in vitamin B6, did not attenuate our estimates suggesting that food rich in vitamin B6 are unlikely to explain our results.

A second explanation is that circulating vitamin B6 varies significantly between cases and controls not due to differences in food intake but from diverse biological processes including absorption or catabolism of the circulating micronutrients. This could induce lower concentrations among cases participants even if their food consumption was similar to controls. Accordingly, results of this would mean that modification through diet might not be a good candidate for reducing cancer incidence. If it turns out that relations with vitamin B6 are causal, it would be very important to identify which of these two interpretations is true.

When we were able to capture the relation of food intake rich on B-vitamins in the HNC and RCC studies (Chapter III and IV), data provided by the LC3 cohorts was incomplete and did not permit us to control for such exposure. Also, we had partial information on supplementation used (dietary or vitamins) that is very common among US population<sup>270</sup>. It is plausible that in the LC3 study, US participants with the highest circulating of B-vitamins tend to be healthier with important use supplement minerals and vitamins. Numerous studies have shown that supplement use is associated with a healthier lifestyle. For example, it was reported that supplement users compared to non-users have higher education and incomes and are more likely to be physically active. In contrast, they are less likely to be tobacco or alcohol consumers compared to non-users<sup>271</sup>. Because (1) the design of LC3 study is prospective, (2) cases and controls were carefully matched on sex, age, cohort and smoking status, and (3) we adjusted by well-known risk factors of lung cancer, confounding effect may be reduced and any bias would be non-differential. However, we could not assure adjustment and cannot exclude the possibility that this missing information on dietary intake and vitamin supplementation affected our estimates in the LC3 study.

## **Main strengths and limitations**

This entire dissertation is based on prospective studies, which are known to be the best design in term of observational epidemiology. Indeed, in retrospective case-control studies reversed causality and confounding by treatment are particularly a concern because the exposure is assessed at or after diagnosis of cancer cases which lead to many biases. In HNC, RCC and LC3 studies, cases were enrolled prior to their cancer diagnosis with information on different variables including baseline characteristics variables as well as biomarker measurements. Thus, this minimizes the chance that any differences between cases and controls are caused by existing tumors and are therefore better designed than retrospective case-control studies to investigate cancer etiology.

However, prospective studies are large enterprises and demand consequent efforts in order to obtain large sample size populations. One approach to deal with, it is to conduct a case-control study nested within a cohort which has similar precision of prospective cohorts but costs less. Our group had the great opportunity to conduct two nested case-control studies (HNC and RCC studies) from one of the largest single prospective cohorts worldwide, the EPIC cohort, reaching about 520,000 participants of whom almost 360,000 donated a blood sample. Furthermore, different collaborations and efforts permitted our group to form the biggest cohort consortium of lung cancers to date which gathered twenty nested case-control studies and coordinate over 12,000 prospective blood analyses of biomarkers.

Despite EPIC and LC3 studies have strengths in terms of sample size, methods for recording participants information were not always similar. For example, the EPIC cohort includes 10 European countries, and in some of them, incident cancer case identification and follow-up were based on cancer registers where as in other countries used a combination of methods including health insurance records, cancer and pathology registries, as well as direct contact participants and their next-of-kin. This issue was also observed in the LC3 study (See Supplementary Methods of LC3 study). It is also important to note that some of the sub-populations from EPIC and LC3 tend to be healthier than the overall population, thus limiting generalizability. For instance, the NHS or the PHS studies are composed exclusively

of nurses and physicians respectively, or the UK center in the EPIC cohort is mainly composed of vegetarians.

Furthermore, as indicated in the LC3 study, large differences in folate and vitamin B6 concentrations were observed across the cohorts and continents. This observation underlines the large lifestyle differences inherent within this consortium. While this observation could permit to assess a larger range of exposure values, it might also raise the issue if unknown life-style related factor interacts with the exposure of interest. In other words, an association found in one sub-population (e.g. SWHS study which is composed mainly of Chinese never smoking women) might not be identified in another (e.g. the ATBC study composed of Finish men smokers) because of different exposure to unknown interacting factors.

As indicated in HNC and RCC studies, OCM biomarkers were measured from a single blood sample drawn in adulthood and therefore may be imperfect estimates of underlying historical exposure. Intra-individual variation may still be important through time and thus having only one precise measure of recent exposure could lead to misclassification of true underlying historical exposure. However, because cases and controls were matched on blood draw they should not have differential misclassification bias. Furthermore, in the LC3 study we looked explicitly on the variability, the intra-individual variation of biomarkers across two samples, and repeat testing indicated an intra-class correlation (ICC) greater than 0.6 for both circulating vitamin B6 and folate, which suggests a reliable collinearity. Importantly, biosample measurements were centralized and took place in a single laboratory for all studies, including the LC3 study, thus minimizing systematic inter-laboratory variation. Coefficients of variations within and between batches were low for all OCM biomarkers measured across our studies.



## Future directions

Findings from the three studies in this dissertation expanded our scientific knowledge on the potential role of OCM biomarkers in smoking-related cancer etiology and outcome. However, it is difficult to disentangle the role of each micronutrient independently of the other in relation to the outcome because the one-carbon pathway is a complex mechanism in which many vitamins and related metabolites interact together. In line with other studies<sup>13</sup>, we observed for example that circulating folate increases circulation of both vitamin B2 and B6, or lower the folate levels increase the concentrations of homocysteine. It won't be surprising that these biomarkers might be correlated with some other OCM analytes. In order to better understand the interactions between OCM nutrients, one could look at additional cornerstones of the OCM pathway such as glycine or serine which were recently found necessary and sufficient for cell transformation and malignancy<sup>133</sup>. In line with this hypothesis, I recently generated some preliminary analyses using the same datasets and observed interesting results indicating that circulating concentrations of both serine and glycine were significantly and inversely associated to HNC risk.

This thesis is composed of three studies with different cancer sites and suggests different associations. First, the HNC study indicates potential role of homocysteine and folate with cancer of the head and neck, although the results are preliminary and hence it is difficult to make any conclusion. These novel findings need to be replicated in other sufficiently-powered prospective studies as well as in different ethnic populations.

Second, the RCC study strongly suggests that circulating concentrations of vitamin B6 are inversely related to both RCC risk and mortality. In addition to the RCC study, an independent new study from our group (Muller DC *et al.*, personal communication) strongly supports the survival benefit of vitamin B6 after diagnosis of RCC. Even though our findings are strong and consistent they need to be replicated in other future large-scale prospective studies as well as in some other ethnic groups than Western. One could envisage in a near future to gather studies across the globe as it was done within lung cancer cohort consortium. This will contribute to our understanding of the role of circulating vitamin B6 in cancer risk and survival, and its exact role in the pathogenesis of this cancer. However, even if our observations regarding circulating vitamin B6 are shown to be causal, it might not be straightforward to confirm that OCM pathway is implicated in this relation. Indeed, multiple

pathways implicated in cancer development are dependent on vitamin B6, examples of which include the tryptophan metabolism pathway, which is involved in immune function and inflammatory processes<sup>272-274</sup>, as well as cell proliferation, angiogenesis, and oxidative stress<sup>149</sup>, the latter being particularly relevant in smoking-related cancers. For example, inflammation can be the underlying cause of RCC etiology and outcome whereas circulating vitamin B6 is only an indicator of inflammation and thus, have no direct and independent role in carcinogenesis and survival.

RCTs could help to elucidate these interpretations. However, it is unlikely that intervention trials of vitamin B6 with RCC would be advisable in a very near future because (1) we do not have enough evidence to date to potentially induce an harm to the patient, (2) recent pooled analysis of two randomized trials reported no benefits for all cancer combined among participants randomized to receive vitamin B6<sup>196</sup>, and (3) a large number of study participants would be required to have a relevant statistical power due to the complexity to evaluate impact of vitamin B6 separately controlling for the other nutrient status or/and genetic variation.

One promising alternative which could help in elucidating the role of vitamin B6 in RCC without going through an intervention study is the Mendelian Randomization (MR) approach. This method aims to identify gene variants associated independently with exposure (i.e. with B6 vitamin concentration), and test whether there is an association between any such gene variants and disease status. Given that allocation of gene variants at meiosis is largely at random, and that such variants should not be associated with lifestyle factors, an independent association between a gene variant and disease status should only occur if it is a true causal association and would therefore provide strong evidence of causality. Conclusive examples of this method include the role of cholesterol for coronary heart disease<sup>275</sup> or folate concentrations in mothers and neural tube defects in offspring<sup>276</sup>. Then, identifying further genetic variants that influence circulating vitamin B6 through large collaborative studies and investigate their relation to RCC risk will be also a major topic in the near future.

Third, although findings from LC3 study did not support the strong inverse association between higher circulating vitamin B6 and decreased risk lung cancer, as previously reported in the EPIC lung study<sup>13</sup>, circulating vitamin B6 may contribute in risk prediction of lung cancer. One of the final analysis in the LC3 study indicated that circulating vitamin B6

displays a compatible association across all the cohort consortium with that previously seen in two European cohorts (ATBC and EPIC)<sup>13 185</sup> in current smokers diagnosed close to blood draw, with a 2-fold risk increase for lung cancer happening within 3 years for participants who we defined as vitamin B6 deficient. This observation might support circulating vitamin B6 as a potential useful lung cancer risk predictor close to blood draw.

The rationale for using biomarker is to measure more adequately the biologically relevant exposure. Most biomarkers from both prospective and retrospective studies rely on a single serum or plasma sample, which only provides an estimate of biomarker exposure over the previous days. As it was evoked previously, one approach to assess temporal changes and to better estimate long-life exposure is to study samples from each subject obtained at several time points. However, given financial and logistic issues this won't be possible for all participants of studies, in particular for a cohort. A compromise is to repeat sample measurements for a representative and large enough fraction of the original cohort, as we did in the LC3 study (Chapter IV). Alternatively, one could think on a single blood sample which gives a good estimation of a medium- long-term nutrient status. For example, blood cell folate concentrations respond slowly to changes in folate intake because the erythrocytes, which have a 120-day lifespan, accumulate folate only during erythropoiesis<sup>277</sup>. Biomarkers represent to date an important tool in cancer epidemiology and, advances technologies are rapid and will facilitate the application of the markers to large sample size studies.

In conclusion, one-carbon metabolism is a complicated pathway, and there are a number of gaps in our understanding its relation to cancer, in particular smoking-related cancers. Findings from this thesis research contributed evidence to fill these gaps. This dissertation is a first "pillar" in my long research aim to better understand the OCM implication in the etiology of smoking-related cancers.

## REFERENCES

1. IARC, editor. *Tobacco smoke and involuntary smoking/ IARC Working Group on the Evaluation of Carcinogenic Risks to Humans (2002: Lyon, France)*. Lyon: International Agency for Research on Cancer, 2004.
2. Jha P. Avoidable global cancer deaths and total deaths from smoking. *Nature reviews. Cancer* 2009;9(9):655-64.
3. Doll R, Peto R, Boreham J, Sutherland I. Mortality from cancer in relation to smoking: 50 years observations on British doctors. *British journal of cancer* 2005;92(3):426-9.
4. Cancer IAFRo. *World Cancer Report*. BERNARD W. STEWART and CHRISTOPHER P. WILD ed. Lyon: Published by the International Agency for Research on Cancer, 2014.
5. Limeback H. *Comprehensive Preventive Dentistry*. USA: John Wiley & Sons, 2012.
6. Trichopoulos D, Bamia C, Lagiou P, Fedirko V, Trepo E, Jenab M, et al. Hepatocellular carcinoma risk factors and disease burden in a European cohort: a nested case-control study. *J Natl Cancer Inst* 2011;103(22):1686-95.
7. Hunt JD, van der Hel OL, McMillan GP, Boffetta P, Brennan P. Renal cell carcinoma in relation to cigarette smoking: meta-analysis of 24 studies. *International journal of cancer. Journal international du cancer* 2005;114(1):101-8.
8. Peto R, Darby S, Deo H, Silcocks P, Whitley E, Doll R. Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. *BMJ* 2000;321(7257):323-9.
9. Howlander N NA, Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975-2012. *National Cancer Institute. Bethesda* 2014.
10. Donaldson MS. Nutrition and cancer: a review of the evidence for an anti-cancer diet. *Nutrition journal* 2004;3:19.
11. Key TJ, Schatzkin A, Willett WC, Allen NE, Spencer EA, Travis RC. Diet, nutrition and the prevention of cancer. *Public health nutrition* 2004;7(1A):187-200.
12. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public health nutrition* 2002;5(6B):1113-24.
13. Johansson M, Relton C, Ueland PM, Vollset SE, Midttun O, Nygard O, et al. Serum B vitamin levels and risk of lung cancer. *Jama* 2010;303(23):2377-85.
14. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA: a cancer journal for clinicians* 2011;61(2):69-90.
15. Ferlay J SI, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. *Lyon, France: International Agency for Research on Cancer* 2013;GLOBOCAN 2012 v1.0.
16. Harris JE. Cigarette smoking among successive birth cohorts of men and women in the United States during 1900-80. *J Natl Cancer Inst* 1983;71(3):473-9.
17. Bray FI, Weiderpass E. Lung cancer mortality trends in 36 European countries: secular trends and birth cohort patterns by sex and region 1970-2007. *International journal of cancer. Journal international du cancer* 2010;126(6):1454-66.
18. Jemal A, Thun MJ, Ries LA, Howe HL, Weir HK, Center MM, et al. Annual report to the nation on the status of cancer, 1975-2005, featuring trends in lung cancer, tobacco use, and tobacco control. *J Natl Cancer Inst* 2008;100(23):1672-94.
19. Peto R LA, Boreham J, Thun M. Mortality From Smoking in Developed Countries 1950-2000. 2006;2nd ed. Revised June 2006.
20. Altekruse SF KC, Krapcho M, Neyman N, Aminou R, Waldron W, Ruhl J, Howlander N, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Cronin K, Chen HS, Feuer EJ, Stinchcomb DG, Edwards BK (eds). . SEER Cancer Statistics Review, 1975-2007. *National Cancer Institute. Bethesda* 2010.

21. Sant M, Allemani C, Santaquilani M, Knijn A, Marchesi F, Capocaccia R. EURO CARE-4. Survival of cancer patients diagnosed in 1995-1999. Results and commentary. *European journal of cancer* 2009;45(6):931-91.
22. Abidin AZ, Garassino MC, Califano R, Harle A, Blackhall F. Targeted therapies in small cell lung cancer: a review. *Therapeutic advances in medical oncology* 2010;2(1):25-37.
23. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clinic proceedings* 2008;83(5):584-94.
24. Levin ML, Goldstein H, Gerhardt PR. Cancer and tobacco smoking; a preliminary report. *Journal of the American Medical Association* 1950;143(4):336-8.
25. Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. *British medical journal* 1950;2(4682):739-48.
26. Doll R, Hill AB. Mortality in Relation to Smoking: Ten Years' Observations of British Doctors. *British medical journal* 1964;1(5396):1460-7 CONCL.
27. Cancer IAFRo, editor. *Tobacco smoking. IARC monograph on the Evaluation of the Carcinogenic Risk of Chemicals to Humans*. Lyon: International Agency for Research on Cancer, 1986.
28. Doll R, Peto R. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. *Journal of epidemiology and community health* 1978;32(4):303-13.
29. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ* 2004;328(7455):1519.
30. Thun MJ, Hannan LM, Adams-Campbell LL, Boffetta P, Buring JE, Feskanich D, et al. Lung cancer occurrence in never-smokers: an analysis of 13 cohorts and 22 cancer registry studies. *PLoS medicine* 2008;5(9):e185.
31. Mu L, Liu L, Niu R, Zhao B, Shi J, Li Y, et al. Indoor air pollution and risk of lung cancer among Chinese female non-smokers. *Cancer causes & control : CCC* 2013;24(3):439-50.
32. Subramanian J, Govindan R. Lung cancer in never smokers: a review. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2007;25(5):561-70.
33. Samet JM, Avila-Tang E, Boffetta P, Hannan LM, Olivo-Marston S, Thun MJ, et al. Lung cancer in never smokers: clinical epidemiology and environmental risk factors. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2009;15(18):5626-45.
34. Darby S, Hill D, Auvinen A, Barros-Dios JM, Baysson H, Bochicchio F, et al. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *BMJ* 2005;330(7485):223.
35. IARC, editor. *IONIZING RADIATION, PART 1: X- AND GAMMA ( $\gamma$ )-RADIATION, AND NEUTRONS. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Lyon: International Agency for Research on Cancer, 2000.
36. Chiou HY, Hsueh YM, Liaw KF, Horng SF, Chiang MH, Pu YS, et al. Incidence of internal cancers and ingested inorganic arsenic: a seven-year follow-up study in Taiwan. *Cancer research* 1995;55(6):1296-300.
37. Research WCRFAIfC, editor. *Food Nutrition and the Prevention of Cancer: a global perspective*. Washington, DC: World Cancer Research Fund, 2007.
38. IARC, editor. *Alcohol consumption and ethyl carbamate/ IARC Working Group on the Evaluation of Carcinogenic Risks to Humans (2007: Lyon, France)*. Lyon: International Agency for Research on Cancer, 2010.
39. Bagnardi V, Rota M, Botteri E, Tramacere I, Islami F, Fedirko V, et al. Light alcohol drinking and cancer: a meta-analysis. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 2013;24(2):301-8.
40. Wang M, Qin S, Zhang T, Song X, Zhang S. The effect of fruit and vegetable intake on the development of lung cancer: a meta-analysis of 32 publications and 20 414 cases. *European journal of clinical nutrition* 2015.
41. IARC, editor. *IARC Handbooks of Cancer Prevention. Carotenoids*. Lyon: International Agency For Research On Cancer, 1998.
42. Ruano-Ravina A, Figueiras A, Freire-Garabal M, Barros-Dios JM. Antioxidant vitamins and risk of lung cancer. *Current pharmaceutical design* 2006;12(5):599-613.

43. Carvalho AL, Nishimoto IN, Califano JA, Kowalski LP. Trends in incidence and prognosis for head and neck cancer in the United States: a site-specific analysis of the SEER database. *International journal of cancer. Journal internationale du cancer* 2005;114(5):806-16.
44. Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. *CA: a cancer journal for clinicians* 2007;57(1):43-66.
45. Schottenfeld D FJ, editor. *Cancer Epidemiology and Prevention*: Oxford, 2006.
46. Argiris A, Karamouzis MV, Raben D, Ferris RL. Head and neck cancer. *Lancet* 2008;371(9625):1695-709.
47. Levin P BaB, editor. *World Cancer Report 2008*. Lyon: International Agency for Research on Cancer (IARC), 2008.
48. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA: a cancer journal for clinicians* 2005;55(2):74-108.
49. Pulte D, Brenner H. Changes in Survival in Head and Neck Cancers in the Late 20th and Early 21st Century: A Period Analysis. *Oncologist* 2010;15(9):994-1001.
50. Olshan AF. *Epidemiology, Pathogenesis, and Prevention of the Head and Neck Cancer*. New-York: Springer, 2010.
51. Gandini S, Botteri E, Iodice S, Boniol M, Lowenfels AB, Maisonneuve P, et al. Tobacco smoking and cancer: a meta-analysis. *International journal of cancer. Journal internationale du cancer* 2008;122(1):155-64.
52. Lubin JH, Purdue M, Kelsey K, Zhang ZF, Winn D, Wei Q, et al. Total exposure and exposure rate effects for alcohol and smoking and risk of head and neck cancer: a pooled analysis of case-control studies. *Am J Epidemiol* 2009;170(8):937-47.
53. Zhang Y, Wang R, Miao L, Zhu L, Jiang H, Yuan H. Different levels in alcohol and tobacco consumption in head and neck cancer patients from 1957 to 2013. *PloS one* 2015;10(4):e0124045.
54. Maasland DH, van den Brandt PA, Kremer B, Goldbohm RA, Schouten LJ. Alcohol consumption, cigarette smoking and the risk of subtypes of head-neck cancer: results from the Netherlands Cohort Study. *BMC cancer* 2014;14:187.
55. Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, Bouvard V, et al. Carcinogenicity of alcoholic beverages. *The Lancet. Oncology* 2007;8(4):292-3.
56. Hashibe M, Brennan P, Benhamou S, Castellsague X, Chen C, Curado MP, et al. Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *J Natl Cancer Inst* 2007;99(10):777-89.
57. Old LJ, Boyse EA, Oettgen HF, Harven ED, Geering G, Williamson B, et al. Precipitating antibody in human serum to an antigen present in cultured burkitt's lymphoma cells. *Proceedings of the National Academy of Sciences of the United States of America* 1966;56(6):1699-704.
58. Mehanna H, Jones TM, Gregoire V, Ang KK. Oropharyngeal carcinoma related to human papillomavirus. *BMJ* 2010;340:c1439.
59. Kreimer AR, Johansson M, Waterboer T, Kaaks R, Chang-Claude J, Drogen D, et al. Evaluation of human papillomavirus antibodies and risk of subsequent head and neck cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2013;31(21):2708-15.
60. Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst* 1981;66(6):1191-308.
61. Boeing H, Dietrich T, Hoffmann K, Pischon T, Ferrari P, Lahmann PH, et al. Intake of fruits and vegetables and risk of cancer of the upper aero-digestive tract: the prospective EPIC-study. *Cancer causes & control : CCC* 2006;17(7):957-69.
62. Chuang SC, Jenab M, Heck JE, Bosetti C, Talamini R, Matsuo K, et al. Diet and the risk of head and neck cancer: a pooled analysis in the INHANCE consortium. *Cancer causes & control : CCC* 2012;23(1):69-88.
63. Aune D, De Stefani E, Ronco A, Boffetta P, Deneo-Pellegrini H, Acosta G, et al. Meat consumption and cancer risk: a case-control study in Uruguay. *Asian Pacific journal of cancer prevention : APJCP* 2009;10(3):429-36.

64. La Vecchia C, Chatenoud L, Franceschi S, Soler M, Parazzini F, Negri E. Vegetables and fruit and human cancer: update of an Italian study. *International journal of cancer. Journal international du cancer* 1999;82(1):151-2.
65. Peters ES, Lockett BG, Applebaum KM, Marsit CJ, McClean MD, Kelsey KT. Dairy products, leanness, and head and neck squamous cell carcinoma. *Head & neck* 2008;30(9):1193-205.
66. Sapkota A, Hsu CC, Zaridze D, Shangina O, Szeszenia-Dabrowska N, Mates D, et al. Dietary risk factors for squamous cell carcinoma of the upper aerodigestive tract in central and eastern Europe. *Cancer causes & control : CCC* 2008;19(10):1161-70.
67. Gupta PC, Hebert JR, Bhonsle RB, Sinor PN, Mehta H, Mehta FS. Dietary factors in oral leukoplakia and submucous fibrosis in a population-based case control study in Gujarat, India. *Oral diseases* 1998;4(3):200-6.
68. Edefonti V, Hashibe M, Parpinel M, Ferraroni M, Turati F, Serraino D, et al. Vitamin E intake from natural sources and head and neck cancer risk: a pooled analysis in the International Head and Neck Cancer Epidemiology consortium. *British journal of cancer* 2015;113(1):182-92.
69. Negri E, Franceschi S, Bosetti C, Levi F, Conti E, Parpinel M, et al. Selected micronutrients and oral and pharyngeal cancer. *International journal of cancer. Journal international du cancer* 2000;86(1):122-7.
70. Galeone C, Edefonti V, Parpinel M, Leoncini E, Matsuo K, Talamini R, et al. Folate intake and the risk of oral cavity and pharyngeal cancer: a pooled analysis within the International Head and Neck Cancer Epidemiology Consortium. *International journal of cancer. Journal international du cancer* 2015;136(4):904-14.
71. Arnold M, Soerjomataram I, Ferlay J, Forman D. Global incidence of oesophageal cancer by histological subtype in 2012. *Gut* 2015;64(3):381-7.
72. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA: a cancer journal for clinicians* 2014;64(1):9-29.
73. Castro C, Bosetti C, Malvezzi M, Bertuccio P, Levi F, Negri E, et al. Patterns and trends in esophageal cancer mortality and incidence in Europe (1980-2011) and predictions to 2015. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 2014;25(1):283-90.
74. Cummings LC, Cooper GS. Descriptive epidemiology of esophageal carcinoma in the Ohio Cancer Registry. *Cancer detection and prevention* 2008;32(1):87-92.
75. De Angelis R, Sant M, Coleman MP, Francisci S, Baili P, Pierannunzio D, et al. Cancer survival in Europe 1999-2007 by country and age: results of EURO CARE--5-a population-based study. *The Lancet. Oncology* 2014;15(1):23-34.
76. Freedman ND, Abnet CC, Leitzmann MF, Mouw T, Subar AF, Hollenbeck AR, et al. A prospective study of tobacco, alcohol, and the risk of esophageal and gastric cancer subtypes. *Am J Epidemiol* 2007;165(12):1424-33.
77. Mao WM, Zheng WH, Ling ZQ. Epidemiologic risk factors for esophageal cancer development. *Asian Pacific journal of cancer prevention : APJCP* 2011;12(10):2461-6.
78. Morita M, Kumashiro R, Kubo N, Nakashima Y, Yoshida R, Yoshinaga K, et al. Alcohol drinking, cigarette smoking, and the development of squamous cell carcinoma of the esophagus: epidemiology, clinical findings, and prevention. *International journal of clinical oncology* 2010;15(2):126-34.
79. Steevens J, Schouten LJ, Goldbohm RA, van den Brandt PA. Alcohol consumption, cigarette smoking and risk of subtypes of oesophageal and gastric cancer: a prospective cohort study. *Gut* 2010;59(1):39-48.
80. Zhang Y. Epidemiology of esophageal cancer. *World journal of gastroenterology : WJG* 2013;19(34):5598-606.
81. Esophageal cancer: epidemiology, pathogenesis and prevention. *Nature clinical practice. Gastroenterology & hepatology* 2008;5(9):517-26.
82. Lee CH, Lee JM, Wu DC, Hsu HK, Kao EL, Huang HL, et al. Independent and combined effects of alcohol intake, tobacco smoking and betel quid chewing on the risk of esophageal cancer in Taiwan. *International journal of cancer. Journal international du cancer* 2005;113(3):475-82.
83. Phukan RK, Ali MS, Chetia CK, Mahanta J. Betel nut and tobacco chewing; potential risk factors of cancer of oesophagus in Assam, India. *British journal of cancer* 2001;85(5):661-7.
84. Wu IC, Lu CY, Kuo FC, Tsai SM, Lee KW, Kuo WR, et al. Interaction between cigarette, alcohol and betel nut use on esophageal cancer risk in Taiwan. *European journal of clinical investigation* 2006;36(4):236-41.

85. Sakata K, Hoshiyama Y, Morioka S, Hashimoto T, Takeshita T, Tamakoshi A. Smoking, alcohol drinking and esophageal cancer: findings from the JACC Study. *Journal of epidemiology / Japan Epidemiological Association* 2005;15 Suppl 2:S212-9.
86. Islami F, Fedirko V, Tramacere I, Bagnardi V, Jenab M, Scotti L, et al. Alcohol drinking and esophageal squamous cell carcinoma with focus on light-drinkers and never-smokers: a systematic review and meta-analysis. *International journal of cancer. Journal international du cancer* 2011;129(10):2473-84.
87. Prabhu A, Obi KO, Rubenstein JH. The synergistic effects of alcohol and tobacco consumption on the risk of esophageal squamous cell carcinoma: a meta-analysis. *The American journal of gastroenterology* 2014;109(6):822-7.
88. Cancer IAfRo. *IARC Handbooks of Cancer Prevention, Fruits and Vegetables*. Lyon: IARC Press 2003.
89. Hajizadeh B, Jessri M, Moasheri SM, Rad AH, Rashidkhani B. Fruits and vegetables consumption and esophageal squamous cell carcinoma: a case-control study. *Nutr Cancer* 2011;63(5):707-13.
90. Liu J, Wang J, Leng Y, Lv C. Intake of fruit and vegetables and risk of esophageal squamous cell carcinoma: a meta-analysis of observational studies. *International journal of cancer. Journal international du cancer* 2013;133(2):473-85.
91. Yamaji T, Inoue M, Sasazuki S, Iwasaki M, Kurahashi N, Shimazu T, et al. Fruit and vegetable consumption and squamous cell carcinoma of the esophagus in Japan: the JPHC study. *International journal of cancer. Journal international du cancer* 2008;123(8):1935-40.
92. Aune D, Deneo-Pellegrini H, Ronco AL, Boffetta P, Acosta G, Mendilaharsu M, et al. Dietary folate intake and the risk of 11 types of cancer: a case-control study in Uruguay. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 2011;22(2):444-51.
93. Galeone C, Pelucchi C, Levi F, Negri E, Talamini R, Franceschi S, et al. Folate intake and squamous-cell carcinoma of the oesophagus in Italian and Swiss men. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 2006;17(3):521-5.
94. Larsson SC, Giovannucci E, Wolk A. Folate intake, MTHFR polymorphisms, and risk of esophageal, gastric, and pancreatic cancer: a meta-analysis. *Gastroenterology* 2006;131(4):1271-83.
95. La Vecchia C, Altieri A, Tavani A. Vegetables, fruit, antioxidants and cancer: a review of Italian studies. *European journal of nutrition* 2001;40(6):261-7.
96. Xiao Q, Freedman ND, Ren J, Hollenbeck AR, Abnet CC, Park Y. Intakes of folate, methionine, vitamin B6, and vitamin B12 with risk of esophageal and gastric cancer in a large cohort study. *British journal of cancer* 2014;110(5):1328-33.
97. Antonsson A, Nancarrow DJ, Brown IS, Green AC, Drew PA, Watson DI, et al. High-risk human papillomavirus in esophageal squamous cell carcinoma. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2010;19(8):2080-7.
98. Cao F, Han H, Zhang F, Wang B, Ma W, Wang Y, et al. HPV infection in esophageal squamous cell carcinoma and its relationship to the prognosis of patients in northern China. *TheScientificWorldJournal* 2014;2014:804738.
99. Chang F, Syrjanen S, Shen Q, Cintorino M, Santopietro R, Tosi P, et al. Evaluation of HPV, CMV, HSV and EBV in esophageal squamous cell carcinomas from a high-incidence area of China. *Anticancer research* 2000;20(5C):3935-40.
100. Syrjanen KJ. HPV infections and oesophageal cancer. *Journal of clinical pathology* 2002;55(10):721-8.
101. Guo F, Liu Y, Wang X, He Z, Weiss NS, Madeleine MM, et al. Human papillomavirus infection and esophageal squamous cell carcinoma: a case-control study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2012;21(5):780-5.
102. Samanic C, Chow WH, Gridley G, Jarvholm B, Fraumeni JF, Jr. Relation of body mass index to cancer risk in 362,552 Swedish men. *Cancer causes & control : CCC* 2006;17(7):901-9.
103. Lahmann PH, Pandeya N, Webb PM, Green AC, Whitman DC. Body mass index, long-term weight change, and esophageal squamous cell carcinoma: is the inverse association modified by smoking status? *Cancer* 2012;118(7):1901-9.



104. Chow WH, Dong LM, Devesa SS. Epidemiology and risk factors for kidney cancer. *Nature reviews. Urology* 2010;7(5):245-57.
105. Patard JJ, Leray E, Rioux-Leclercq N, Cindolo L, Ficarra V, Zisman A, et al. Prognostic value of histologic subtypes in renal cell carcinoma: a multicenter experience. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2005;23(12):2763-71.
106. Janout V, Janoutova G. Epidemiology and risk factors of kidney cancer. *Biomedical papers of the Medical Faculty of the University Palacky, Olomouc, Czechoslovakia* 2004;148(1):95-101.
107. Yuan JM, Castela JE, Gago-Dominguez M, Yu MC, Ross RK. Tobacco use in relation to renal cell carcinoma. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 1998;7(5):429-33.
108. Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nature reviews. Cancer* 2004;4(8):579-91.
109. Adams KF, Leitzmann MF, Albanes D, Kipnis V, Moore SC, Schatzkin A, et al. Body size and renal cell cancer incidence in a large US cohort study. *Am J Epidemiol* 2008;168(3):268-77.
110. Pischon T, Lahmann PH, Boeing H, Tjonneland A, Halkjaer J, Overvad K, et al. Body size and risk of renal cell carcinoma in the European Prospective Investigation into Cancer and Nutrition (EPIC). *International journal of cancer. Journal international du cancer* 2006;118(3):728-38.
111. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. *BMJ* 2007;335(7630):1134.
112. Stojanovic M, Goldner B, Ivkovic D. Renal cell carcinoma and arterial hypertension. *Clinical and experimental nephrology* 2009;13(4):295-9.
113. Choi MY, Jee SH, Sull JW, Nam CM. The effect of hypertension on the risk for kidney cancer in Korean men. *Kidney international* 2005;67(2):647-52.
114. Chow WH, Gridley G, Fraumeni JF, Jr., Jarvholm B. Obesity, hypertension, and the risk of kidney cancer in men. *The New England journal of medicine* 2000;343(18):1305-11.
115. Vatten LJ, Trichopoulos D, Holmen J, Nilsen TI. Blood pressure and renal cancer risk: the HUNT Study in Norway. *British journal of cancer* 2007;97(1):112-4.
116. Weikert S, Boeing H, Pischon T, Weikert C, Olsen A, Tjonneland A, et al. Blood pressure and risk of renal cell carcinoma in the European prospective investigation into cancer and nutrition. *Am J Epidemiol* 2008;167(4):438-46.
117. Flaherty KT, Fuchs CS, Colditz GA, Stampfer MJ, Speizer FE, Willett WC, et al. A prospective study of body mass index, hypertension, and smoking and the risk of renal cell carcinoma (United States). *Cancer causes & control : CCC* 2005;16(9):1099-106.
118. Zucchetto A, Dal Maso L, Tavani A, Montella M, Ramazzotti V, Talamini R, et al. History of treated hypertension and diabetes mellitus and risk of renal cell cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 2007;18(3):596-600.
119. Mattioli S, Truffelli D, Baldasseroni A, Risi A, Marchesini B, Giacomini C, et al. Occupational risk factors for renal cell cancer: a case-control study in northern Italy. *Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine* 2002;44(11):1028-36.
120. Rashidkhani B, Akesson A, Lindblad P, Wolk A. Alcohol consumption and risk of renal cell carcinoma: a prospective study of Swedish women. *International journal of cancer. Journal international du cancer* 2005;117(5):848-53.
121. Mucci LA, Lindblad P, Steineck G, Adami HO. Dietary acrylamide and risk of renal cell cancer. *International journal of cancer. Journal international du cancer* 2004;109(5):774-6.
122. Wozniak MB, Brennan P, Brenner DR, Overvad K, Olsen A, Tjonneland A, et al. Alcohol consumption and the risk of renal cancers in the European prospective investigation into cancer and nutrition (EPIC). *International journal of cancer. Journal international du cancer* 2015;137(8):1953-66.
123. Cheng G, Xie L. Alcohol intake and risk of renal cell carcinoma: a meta-analysis of published case-control studies. *Archives of medical science : AMS* 2011;7(4):648-57.
124. Lee JE, Hunter DJ, Spiegelman D, Adami HO, Albanes D, Bernstein L, et al. Alcohol intake and renal cell cancer in a pooled analysis of 12 prospective studies. *J Natl Cancer Inst* 2007;99(10):801-10.

125. Lee JE, Mannisto S, Spiegelman D, Hunter DJ, Bernstein L, van den Brandt PA, et al. Intakes of fruit, vegetables, and carotenoids and renal cell cancer risk: a pooled analysis of 13 prospective studies. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2009;18(6):1730-9.
126. Wolk A, Lindblad P, Adami HO. Nutrition and renal cell cancer. *Cancer causes & control : CCC* 1996;7(1):5-18.
127. Yuan JM, Gago-Dominguez M, Castela JE, Hankin JH, Ross RK, Yu MC. Cruciferous vegetables in relation to renal cell carcinoma. *International journal of cancer. Journal international du cancer* 1998;77(2):211-6.
128. Hu J, Mao Y, White K. Diet and vitamin or mineral supplements and risk of renal cell carcinoma in Canada. *Cancer causes & control : CCC* 2003;14(8):705-14.
129. Nicodemus KK, Sweeney C, Folsom AR. Evaluation of dietary, medical and lifestyle risk factors for incident kidney cancer in postmenopausal women. *International journal of cancer. Journal international du cancer* 2004;108(1):115-21.
130. Supplements NOoD. Dietary Supplement Fact Sheets.
131. Kim YI. Folate and DNA methylation: a mechanistic link between folate deficiency and colorectal cancer? *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2004;13(4):511-9.
132. Kim YI. Nutritional epigenetics: impact of folate deficiency on DNA methylation and colon cancer susceptibility. *The Journal of nutrition* 2005;135(11):2703-9.
133. Locasale JW. Serine, glycine and one-carbon units: cancer metabolism in full circle. *Nature reviews. Cancer* 2013;13(8):572-83.
134. Olsen A, Halkjaer J, van Gils CH, Buijsse B, Verhagen H, Jenab M, et al. Dietary intake of the water-soluble vitamins B1, B2, B6, B12 and C in 10 countries in the European Prospective Investigation into Cancer and Nutrition. *European journal of clinical nutrition* 2009;63 Suppl 4:S122-49.
135. U ZHaT, editor. *DNA Methylation and Cancer In: Epigenetics and Cancer*. Oxford, UK: ELSEVIER, 2010.
136. Fukushima M, Morita M, Ikeda K, Nagayama S. Population study of expression of thymidylate synthase and dihydropyrimidine dehydrogenase in patients with solid tumors. *International journal of molecular medicine* 2003;12(6):839-44.
137. Hutchinson F. Use of data from bacteria to interpret data on DNA damage processing in mammalian cells. *Mutation research* 1989;220(2-3):269-78.
138. Blount BC, Mack MM, Wehr CM, MacGregor JT, Hiatt RA, Wang G, et al. Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage. *Proceedings of the National Academy of Sciences of the United States of America* 1997;94(7):3290-5.
139. Blount BC, Ames BN. DNA damage in folate deficiency. *Bailliere's clinical haematology* 1995;8(3):461-78.
140. Knock E, Deng L, Wu Q, Lawrance AK, Wang XL, Rozen R. Strain differences in mice highlight the role of DNA damage in neoplasia induced by low dietary folate. *The Journal of nutrition* 2008;138(4):653-8.
141. Cravo ML, Mason JB, Dayal Y, Hutchinson M, Smith D, Selhub J, et al. Folate deficiency enhances the development of colonic neoplasia in dimethylhydrazine-treated rats. *Cancer research* 1992;52(18):5002-6.
142. Ulrich CM, Potter JD. Folate and cancer--timing is everything. *Jama* 2007;297(21):2408-9.
143. Ulrich CM, Potter JD. Folate supplementation: too much of a good thing? *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2006;15(2):189-93.
144. DiSorbo DM, Wagner R, Jr., Nathanson L. In vivo and in vitro inhibition of B16 melanoma growth by vitamin B6. *Nutr Cancer* 1985;7(1-2):43-52.
145. Galluzzi L, Vacchelli E, Michels J, Garcia P, Kepp O, Senovilla L, et al. Effects of vitamin B6 metabolism on oncogenesis, tumor progression and therapeutic responses. *Oncogene* 2013;32(42):4995-5004.
146. Molina A, Oka T, Munoz SM, Chikamori-Aoyama M, Kuwahata M, Natori Y. Vitamin B6 suppresses growth and expression of albumin gene in a human hepatoma cell line HepG2. *Nutr Cancer* 1997;28(2):206-11.
147. Larsson SC, Orsini N, Wolk A. Vitamin B6 and risk of colorectal cancer: a meta-analysis of prospective studies. *Jama* 2010;303(11):1077-83.

148. Komatsu SI, Watanabe H, Oka T, Tsuge H, Nii H, Kato N. Vitamin B-6-supplemented diets compared with a low vitamin B-6 diet suppress azoxymethane-induced colon tumorigenesis in mice by reducing cell proliferation. *The Journal of nutrition* 2001;131(8):2204-7.
149. Matsubara K, Komatsu S, Oka T, Kato N. Vitamin B6-mediated suppression of colon tumorigenesis, cell proliferation, and angiogenesis (review). *The Journal of nutritional biochemistry* 2003;14(5):246-50.
150. Shen J, Lai CQ, Mattei J, Ordovas JM, Tucker KL. Association of vitamin B-6 status with inflammation, oxidative stress, and chronic inflammatory conditions: the Boston Puerto Rican Health Study. *The American journal of clinical nutrition* 2010;91(2):337-42.
151. Bird A. DNA methylation patterns and epigenetic memory. *Genes & development* 2002;16(1):6-21.
152. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nature reviews. Genetics* 2002;3(6):415-28.
153. Wolffe AP, Matzke MA. Epigenetics: regulation through repression. *Science* 1999;286(5439):481-6.
154. Paz MF, Fraga MF, Avila S, Guo M, Pollan M, Herman JG, et al. A systematic profile of DNA methylation in human cancer cell lines. *Cancer research* 2003;63(5):1114-21.
155. Das PM, Singal R. DNA methylation and cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2004;22(22):4632-42.
156. Catteau A, Morris JR. BRCA1 methylation: a significant role in tumour development? *Seminars in cancer biology* 2002;12(5):359-71.
157. Szyf M. DNA methylation signatures for breast cancer classification and prognosis. *Genome medicine* 2012;4(3):26.
158. Lee WH, Morton RA, Epstein JI, Brooks JD, Campbell PA, Bova GS, et al. Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 1994;91(24):11733-7.
159. Leone G, Teofili L, Voso MT, Lubbert M. DNA methylation and demethylating drugs in myelodysplastic syndromes and secondary leukemias. *Haematologica* 2002;87(12):1324-41.
160. Toyota M, Kopecky KJ, Toyota MO, Jair KW, Willman CL, Issa JP. Methylation profiling in acute myeloid leukemia. *Blood* 2001;97(9):2823-9.
161. Tsou JA, Hagen JA, Carpenter CL, Laird-Offringa IA. DNA methylation analysis: a powerful new tool for lung cancer diagnosis. *Oncogene* 2002;21(35):5450-61.
162. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 1983;301(5895):89-92.
163. Kim YI, Giuliano A, Hatch KD, Schneider A, Nour MA, Dallal GE, et al. Global DNA hypomethylation increases progressively in cervical dysplasia and carcinoma. *Cancer* 1994;74(3):893-9.
164. Bedford MT, van Helden PD. Hypomethylation of DNA in pathological conditions of the human prostate. *Cancer research* 1987;47(20):5274-6.
165. Lin CH, Hsieh SY, Sheen IS, Lee WC, Chen TC, Shyu WC, et al. Genome-wide hypomethylation in hepatocellular carcinogenesis. *Cancer research* 2001;61(10):4238-43.
166. Ehrlich M. DNA methylation in cancer: too much, but also too little. *Oncogene* 2002;21(35):5400-13.
167. Chalitchagorn K, Shuangshoti S, Hourpai N, Kongruttanachok N, Tangkijvanich P, Thong-ngam D, et al. Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis. *Oncogene* 2004;23(54):8841-6.
168. Goelz SE, Vogelstein B, Hamilton SR, Feinberg AP. Hypomethylation of DNA from benign and malignant human colon neoplasms. *Science* 1985;228(4696):187-90.
169. Suter CM, Martin DI, Ward RL. Hypomethylation of L1 retrotransposons in colorectal cancer and adjacent normal tissue. *International journal of colorectal disease* 2004;19(2):95-101.
170. Shin W, Yan J, Abratte CM, Vermeylen F, Caudill MA. Choline intake exceeding current dietary recommendations preserves markers of cellular methylation in a genetic subgroup of folate-compromised men. *The Journal of nutrition* 2010;140(5):975-80.
171. Yi P, Melnyk S, Pogribna M, Pogribny IP, Hine RJ, James SJ. Increase in plasma homocysteine associated with parallel increases in plasma S-adenosylhomocysteine and lymphocyte DNA hypomethylation. *The Journal of biological chemistry* 2000;275(38):29318-23.

172. Christensen BC, Kelsey KT, Zheng S, Houseman EA, Marsit CJ, Wrensch MR, et al. Breast cancer DNA methylation profiles are associated with tumor size and alcohol and folate intake. *PLoS genetics* 2010;6(7):e1001043.
173. Kim JW, Park HM, Choi YK, Chong SY, Oh D, Kim NK. Polymorphisms in genes involved in folate metabolism and plasma DNA methylation in colorectal cancer patients. *Oncology reports* 2011;25(1):167-72.
174. Hsiung DT, Marsit CJ, Houseman EA, Eddy K, Furniss CS, McClean MD, et al. Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2007;16(1):108-14.
175. Moore LE, Pfeiffer RM, Poscablo C, Real FX, Kogevinas M, Silverman D, et al. Genomic DNA hypomethylation as a biomarker for bladder cancer susceptibility in the Spanish Bladder Cancer Study: a case-control study. *The Lancet. Oncology* 2008;9(4):359-66.
176. de Vogel S, Bongaerts BW, Wouters KA, Kester AD, Schouten LJ, de Goeij AF, et al. Associations of dietary methyl donor intake with MLH1 promoter hypermethylation and related molecular phenotypes in sporadic colorectal cancer. *Carcinogenesis* 2008;29(9):1765-73.
177. Figueiredo JC, Grau MV, Wallace K, Levine AJ, Shen L, Hamdan R, et al. Global DNA hypomethylation (LINE-1) in the normal colon and lifestyle characteristics and dietary and genetic factors. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2009;18(4):1041-9.
178. Kennedy DA, Stern SJ, Moretti M, Matok I, Sarkar M, Nickel C, et al. Folate intake and the risk of colorectal cancer: a systematic review and meta-analysis. *Cancer epidemiology* 2011;35(1):2-10.
179. Kim YI. Folate and carcinogenesis: evidence, mechanisms, and implications. *The Journal of nutritional biochemistry* 1999;10(2):66-88.
180. Kim YI. Folate and colorectal cancer: an evidence-based critical review. *Molecular nutrition & food research* 2007;51(3):267-92.
181. Dai WM, Yang B, Chu XY, Wang YQ, Zhao M, Chen L, et al. Association between folate intake, serum folate levels and the risk of lung cancer: a systematic review and meta-analysis. *Chinese medical journal* 2013;126(10):1957-64.
182. Shen H, Wei Q, Pillow PC, Amos CI, Hong WK, Spitz MR. Dietary folate intake and lung cancer risk in former smokers: a case-control analysis. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2003;12(10):980-6.
183. Zhang YF, Zhou L, Zhang HW, Hou AJ, Gao HF, Zhou YH. Association between folate intake and the risk of lung cancer: a dose-response meta-analysis of prospective studies. *PloS one* 2014;9(4):e93465.
184. Bandera EV, Freudenheim JL, Marshall JR, Zielezny M, Priore RL, Brasure J, et al. Diet and alcohol consumption and lung cancer risk in the New York State Cohort (United States). *Cancer causes & control : CCC* 1997;8(6):828-40.
185. Hartman TJ, Woodson K, Stolzenberg-Solomon R, Virtamo J, Selhub J, Barrett MJ, et al. Association of the B-vitamins pyridoxal 5'-phosphate (B(6)), B(12), and folate with lung cancer risk in older men. *Am J Epidemiol* 2001;153(7):688-94.
186. Takata Y, Xiang YB, Yang G, Li H, Gao J, Cai H, et al. Intakes of fruits, vegetables, and related vitamins and lung cancer risk: results from the Shanghai Men's Health Study (2002-2009). *Nutr Cancer* 2013;65(1):51-61.
187. Takata Y, Cai Q, Beeghly-Fadiel A, Li H, Shrubsole MJ, Ji BT, et al. Dietary B vitamin and methionine intakes and lung cancer risk among female never smokers in China. *Cancer causes & control : CCC* 2012;23(12):1965-75.
188. Bassett JK, Hodge AM, English DR, Baglietto L, Hopper JL, Giles GG, et al. Dietary intake of B vitamins and methionine and risk of lung cancer. *European journal of clinical nutrition* 2012;66(2):182-7.
189. Roswall N, Olsen A, Christensen J, Dragsted LO, Overvad K, Tjønneland A. Source-specific effects of micronutrients in lung cancer prevention. *Lung cancer* 2010;67(3):275-81.

190. Slatore CG, Littman AJ, Au DH, Satia JA, White E. Long-term use of supplemental multivitamins, vitamin C, vitamin E, and folate does not reduce the risk of lung cancer. *American journal of respiratory and critical care medicine* 2008;177(5):524-30.
191. Kabat GC, Miller AB, Jain M, Rohan TE. Dietary intake of selected B vitamins in relation to risk of major cancers in women. *British journal of cancer* 2008;99(5):816-21.
192. Cho E, Hunter DJ, Spiegelman D, Albanes D, Beeson WL, van den Brandt PA, et al. Intakes of vitamins A, C and E and folate and multivitamins and lung cancer: a pooled analysis of 8 prospective studies. *International journal of cancer. Journal international du cancer* 2006;118(4):970-8.
193. Yuan JM, Stram DO, Arakawa K, Lee HP, Yu MC. Dietary cryptoxanthin and reduced risk of lung cancer: the Singapore Chinese Health Study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2003;12(9):890-8.
194. Voorrips LE, Goldbohm RA, Brants HA, van Poppel GA, Sturmans F, Hermus RJ, et al. A prospective cohort study on antioxidant and folate intake and male lung cancer risk. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2000;9(4):357-65.
195. Vollset SE, Clarke R, Lewington S, Ebbing M, Halsey J, Lonn E, et al. Effects of folic acid supplementation on overall and site-specific cancer incidence during the randomised trials: meta-analyses of data on 50,000 individuals. *Lancet* 2013;381(9871):1029-36.
196. Ebbing M, Bonna KH, Nygard O, Arnesen E, Ueland PM, Nordrehaug JE, et al. Cancer incidence and mortality after treatment with folic acid and vitamin B12. *Jama* 2009;302(19):2119-26.
197. Freedman ND, Park Y, Subar AF, Hollenbeck AR, Leitzmann MF, Schatzkin A, et al. Fruit and vegetable intake and head and neck cancer risk in a large United States prospective cohort study. *International journal of cancer. Journal international du cancer* 2008;122(10):2330-6.
198. De Stefani E, Ronco A, Mendilaharsu M, Deneo-Pellegrini H. Diet and risk of cancer of the upper aerodigestive tract--II. Nutrients. *Oral oncology* 1999;35(1):22-6.
199. Weinstein SJ, Gridley G, Harty LC, Diehl SR, Brown LM, Winn DM, et al. Folate intake, serum homocysteine and methylenetetrahydrofolate reductase (MTHFR) C677T genotype are not associated with oral cancer risk in Puerto Rico. *The Journal of nutrition* 2002;132(4):762-7.
200. Bidoli E, Bosetti C, La Vecchia C, Levi F, Parpinel M, Talamini R, et al. Micronutrients and laryngeal cancer risk in Italy and Switzerland: a case-control study. *Cancer causes & control : CCC* 2003;14(5):477-84.
201. Matsuo K, Rossi M, Negri E, Oze I, Hosono S, Ito H, et al. Folate, alcohol, and aldehyde dehydrogenase 2 polymorphism and the risk of oral and pharyngeal cancer in Japanese. *European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation* 2012;21(2):193-8.
202. Pelucchi C, Talamini R, Negri E, Levi F, Conti E, Franceschi S, et al. Folate intake and risk of oral and pharyngeal cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 2003;14(11):1677-81.
203. Almadori G, Bussu F, Galli J, Cadoni G, Zappacosta B, Persichilli S, et al. Serum folate and homocysteine levels in head and neck squamous cell carcinoma. *Cancer* 2002;94(4):1006-11.
204. Almadori G, Bussu F, Galli J, Cadoni G, Zappacosta B, Persichilli S, et al. Serum levels of folate, homocysteine, and vitamin B12 in head and neck squamous cell carcinoma and in laryngeal leukoplakia. *Cancer* 2005;103(2):284-92.
205. Eleftheriadou A, Chalastras T, Ferekidou E, Yiotakis I, Kyriou L, Tzagarakis M, et al. Association between squamous cell carcinoma of the head and neck and serum folate and homocysteine. *Anticancer research* 2006;26(3B):2345-8.
206. Gorgulu O, Selcuk T, Ozdemir S, Sayar C, Beyazit Y, Akbas Y. Evaluation of the roles of serum vitamin B(12), folate and homocysteine levels in laryngeal squamous cell carcinoma. *The Journal of international medical research* 2010;38(6):2047-52.
207. Brown LM, Swanson CA, Gridley G, Swanson GM, Silverman DT, Greenberg RS, et al. Dietary factors and the risk of squamous cell esophageal cancer among black and white men in the United States. *Cancer causes & control : CCC* 1998;9(5):467-74.

208. Mayne ST, Risch HA, Dubrow R, Chow WH, Gammon MD, Vaughan TL, et al. Nutrient intake and risk of subtypes of esophageal and gastric cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2001;10(10):1055-62.
209. Bosetti C, Scotti L, Maso LD, Talamini R, Montella M, Negri E, et al. Micronutrients and the risk of renal cell cancer: a case-control study from Italy. *International journal of cancer. Journal international du cancer* 2007;120(4):892-6.
210. Brock KE, Ke L, Gridley G, Chiu BC, Ershow AG, Lynch CF, et al. Fruit, vegetables, fibre and micronutrients and risk of US renal cell carcinoma. *The British journal of nutrition* 2012;108(6):1077-85.
211. Cho E, Giovannucci EL, Joh HK. Nutrients related to one-carbon metabolism and risk of renal cell cancer. *Cancer causes & control : CCC* 2013;24(2):373-82.
212. Gibson TM, Weinstein SJ, Mayne ST, Pfeiffer RM, Selhub J, Taylor PR, et al. A prospective study of one-carbon metabolism biomarkers and risk of renal cell carcinoma. *Cancer causes & control : CCC* 2010;21(7):1061-9.
213. Wolfgang Ahrens IP. *Handbook of epidemiology*. Berlin Heidelberg: Springer, 2005.
214. JM L. *A dictionary of epidemiology*. 1st edition ed. Oxford: Oxford University Press, 2001.
215. Rothman KJ GS, Lash TL. *Modern epidemiology*. 3rd ed. Philadelphia PA: Lippincott Williams and Wilkins, 2008.
216. Lagiou P, Adami HO, Trichopoulos D. Causality in cancer epidemiology. *European journal of epidemiology* 2005;20(7):565-74.
217. Kovesdy CP, Kalantar-Zadeh K. Observational studies versus randomized controlled trials: avenues to causal inference in nephrology. *Advances in chronic kidney disease* 2012;19(1):11-8.
218. Cole BF, Baron JA, Sandler RS, Haile RW, Ahnen DJ, Bresalier RS, et al. Folic acid for the prevention of colorectal adenomas: a randomized clinical trial. *Jama* 2007;297(21):2351-9.
219. Jaszewski R, Misra S, Tobi M, Ullah N, Naumoff JA, Kucuk O, et al. Folic acid supplementation inhibits recurrence of colorectal adenomas: a randomized chemoprevention trial. *World journal of gastroenterology : WJG* 2008;14(28):4492-8.
220. Logan RF, Grainge MJ, Shepherd VC, Armitage NC, Muir KR. Aspirin and folic acid for the prevention of recurrent colorectal adenomas. *Gastroenterology* 2008;134(1):29-38.
221. Breslow NE DN. Statistical methods in cancer research, Volume II. The design and analysis of cohort studies. *Statistical methods in cancer research*. Lyon: IARC, 1987.
222. Rothman KJ GS. *Modern epidemiology*. 2nd ed. Philadelphia PA: Lippincott Williams and Wilkins, 1998.
223. Dawber TR, Meadors GF, Moore FE, Jr. Epidemiological approaches to heart disease: the Framingham Study. *American journal of public health and the nation's health* 1951;41(3):279-81.
224. Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking; a second report on the mortality of British doctors. *British medical journal* 1956;2(5001):1071-81.
225. Hoffmans MD, Kromhout D, de Lezenne Coulander C. The impact of body mass index of 78,612 18-year old Dutch men on 32-year mortality from all causes. *Journal of clinical epidemiology* 1988;41(8):749-56.
226. Kodama K, Mabuchi K, Shigematsu I. A long-term cohort study of the atomic-bomb survivors. *Journal of epidemiology / Japan Epidemiological Association* 1996;6(3 Suppl):S95-105.
227. I dSS. *Cancer epidemiology: principles and methods*. Lyon: IARC, 1999.
228. Thompson WD. Statistical analysis of case-control studies. *Epidemiologic reviews* 1994;16(1):33-50.
229. Langholz B GL. Risk set sampling in epidemiologic cohort studies. *Statistical Science* 1996;11:35-53.
230. Hans-Olov Adami DH, and Dimitrios Trichopoulos. *Textbook of Cancer Epidemiology*: Oxford Scholarship 2008.
231. Rothman KJ GS. *Validity and Generalizability in Epidemiologic Studies*. In: *Encyclopedia of Biostatistics*: John Wiley & Sons, 2005.

232. Greenland S, Thomas DC. On the need for the rare disease assumption in case-control studies. *Am J Epidemiol* 1982;116(3):547-53.
233. Hak E, Wei F, Grobbee DE, Nichol KL. A nested case-control study of influenza vaccination was a cost-effective alternative to a full cohort analysis. *Journal of clinical epidemiology* 2004;57(9):875-80.
234. Kelsey JL WA, Evans AS, Thompson WD. *Methods in observational epidemiology*. New-York: Oxford University Press, 1996.
235. Slimani N, Kaaks R, Ferrari P, Casagrande C, Clavel-Chapelon F, Lotze G, et al. European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study: rationale, design and population characteristics. *Public health nutrition* 2002;5(6B):1125-45.
236. Ebbing M, Bleie O, Ueland PM, Nordrehaug JE, Nilsen DW, Vollset SE, et al. Mortality and cardiovascular events in patients treated with homocysteine-lowering B vitamins after coronary angiography: a randomized controlled trial. *Jama* 2008;300(7):795-804.
237. Johansson M, Appleby PN, Allen NE, Travis RC, Roddam AW, Egevad L, et al. Circulating concentrations of folate and vitamin B12 in relation to prostate cancer risk: results from the European Prospective Investigation into Cancer and Nutrition study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2008;17(2):279-85.
238. Johansson M, Van Guelpen B, Vollset SE, Hultdin J, Bergh A, Key T, et al. One-carbon metabolism and prostate cancer risk: prospective investigation of seven circulating B vitamins and metabolites. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2009;18(5):1538-43.
239. Midttun O, Hustad S, Ueland PM. Quantitative profiling of biomarkers related to B-vitamin status, tryptophan metabolism and inflammation in human plasma by liquid chromatography/tandem mass spectrometry. *Rapid communications in mass spectrometry : RCM* 2009;23(9):1371-9.
240. Ueland PM, Midttun O, Windelberg A, Svardal A, Skalevik R, Hustad S. Quantitative profiling of folate and one-carbon metabolism in large-scale epidemiological studies by mass spectrometry. *Clinical chemistry and laboratory medicine : CCLM / FESCC* 2007;45(12):1737-45.
241. Molloy AM, Scott JM. Microbiological assay for serum, plasma, and red cell folate using cryopreserved, microtiter plate method. *Methods in enzymology* 1997;281:43-53.
242. Kelleher BP, Broin SD. Microbiological assay for vitamin B12 performed in 96-well microtitre plates. *Journal of clinical pathology* 1991;44(7):592-5.
243. Zhang XH, Ma J, Smith-Warner SA, Lee JE, Giovannucci E. Vitamin B6 and colorectal cancer: current evidence and future directions. *World journal of gastroenterology : WJG* 2013;19(7):1005-10.
244. de Benoist B. Conclusions of a WHO Technical Consultation on folate and vitamin B12 deficiencies. *Food and nutrition bulletin* 2008;29(2 Suppl):S238-44.
245. SAS® 9.4 Help and documentation [program]. 9.4 version. Cary, NC, USA: SAS Institute Inc, 2014.
246. Stata Statistical Software [program]. version 12.1 version, 2011.
247. RStan: the R interface to Stan [program]. Version 2.5 version, 2014.
248. Stan Modeling Language Users Guide and Reference Manual [program]. Version 2.5.0. version, 2014.
249. Brennan P, Crispo A, Zaridze D, Szeszenia-Dabrowska N, Rudnai P, Lissowska J, et al. High cumulative risk of lung cancer death among smokers and nonsmokers in Central and Eastern Europe. *Am J Epidemiol* 2006;164(12):1233-41.
250. Crispo A, Brennan P, Jockel KH, Schaffrath-Rosario A, Wichmann HE, Nyberg F, et al. The cumulative risk of lung cancer among current, ex- and never-smokers in European men. *Br J Cancer* 2004;91(7):1280-6.
251. Ebbert JO, Yang P, Vachon CM, Vierkant RA, Cerhan JR, Folsom AR, et al. Lung cancer risk reduction after smoking cessation: observations from a prospective cohort of women. *J Clin Oncol* 2003;21(5):921-6.
252. Kalmbach RD, Choumenkovitch SF, Troen AM, D'Agostino R, Jacques PF, Selhub J. Circulating folic acid in plasma: relation to folic acid fortification. *Am J Clin Nutr* 2008;88(3):763-8.
253. Boffetta P, Clark S, Shen M, Gislefoss R, Peto R, Andersen A. Serum cotinine level as predictor of lung cancer risk. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2006;15(6):1184-8.

254. Bailey RL, Gahche JJ, Lentino CV, Dwyer JT, Engel JS, Thomas PR, et al. Dietary supplement use in the United States, 2003-2006. *The Journal of nutrition* 2011;141(2):261-6.
255. Jacques PF, Selhub J, Bostom AG, Wilson PW, Rosenberg IH. The effect of folic acid fortification on plasma folate and total homocysteine concentrations. *N Engl J Med* 1999;340(19):1449-54.
256. Dahlin AM, Van Guelpen B, Hultdin J, Johansson I, Hallmans G, Palmqvist R. Plasma vitamin B12 concentrations and the risk of colorectal cancer: a nested case-referent study. *International journal of cancer. Journal international du cancer* 2008;122(9):2057-61.
257. Figueiredo JC, Mott LA, Giovannucci E, Wu K, Cole B, Grainge MJ, et al. Folic acid and prevention of colorectal adenomas: a combined analysis of randomized clinical trials. *International journal of cancer. Journal international du cancer* 2011;129(1):192-203.
258. Le Marchand L, White KK, Nomura AM, Wilkens LR, Selhub JS, Tiirikainen M, et al. Plasma levels of B vitamins and colorectal cancer risk: the multiethnic cohort study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2009;18(8):2195-201.
259. Wei EK, Giovannucci E, Selhub J, Fuchs CS, Hankinson SE, Ma J. Plasma vitamin B6 and the risk of colorectal cancer and adenoma in women. *J Natl Cancer Inst* 2005;97(9):684-92.
260. Chou YC, Lee MS, Wu MH, Shih HL, Yang T, Yu CP, et al. Plasma homocysteine as a metabolic risk factor for breast cancer: findings from a case-control study in Taiwan. *Breast cancer research and treatment* 2007;101(2):199-205.
261. Lin J, Lee IM, Cook NR, Selhub J, Manson JE, Buring JE, et al. Plasma folate, vitamin B-6, vitamin B-12, and risk of breast cancer in women. *The American journal of clinical nutrition* 2008;87(3):734-43.
262. Lin J, Lee IM, Song Y, Cook NR, Selhub J, Manson JE, et al. Plasma homocysteine and cysteine and risk of breast cancer in women. *Cancer research* 2010;70(6):2397-405.
263. Zhang SM, Willett WC, Selhub J, Hunter DJ, Giovannucci EL, Holmes MD, et al. Plasma folate, vitamin B6, vitamin B12, homocysteine, and risk of breast cancer. *J Natl Cancer Inst* 2003;95(5):373-80.
264. Lam WK, White NW, Chan-Yeung MM. Lung cancer epidemiology and risk factors in Asia and Africa. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2004;8(9):1045-57.
265. McCarthy WJ, Meza R, Jeon J, Moolgavkar SH. Chapter 6: Lung cancer in never smokers: epidemiology and risk prediction models. *Risk Anal* 2012;32 Suppl 1:S69-84.
266. Subramanian J, Govindan R. Lung cancer in 'Never-smokers': a unique entity. *Oncology* 2010;24(1):29-35.
267. Rosner B, Gore R. Measurement error correction in nutritional epidemiology based on individual foods, with application to the relation of diet to breast cancer. *Am J Epidemiol* 2001;154(9):827-35.
268. Verkleij-Hagoort AC, de Vries JH, Steegers MP, Lindemans J, Ursem NT, Steegers-Theunissen RP. Validation of the assessment of folate and vitamin B12 intake in women of reproductive age: the method of triads. *European journal of clinical nutrition* 2007;61(5):610-5.
269. Yoshino K, Nishide M, Sankai T, Inagawa M, Yokota K, Moriyama Y, et al. Validity of brief food frequency questionnaire for estimation of dietary intakes of folate, vitamins B6 and B12, and their associations with plasma homocysteine concentrations. *International journal of food sciences and nutrition* 2010;61(1):61-7.
270. Velicer CM, Ulrich CM. Vitamin and mineral supplement use among US adults after cancer diagnosis: a systematic review. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2008;26(4):665-73.
271. Hoggatt KJ. Commentary: Vitamin supplement use and confounding by lifestyle. *International journal of epidemiology* 2003;32(4):553-5.
272. Midttun O, Ulvik A, Ringdal Pedersen E, Ebbing M, Bleie O, Schartum-Hansen H, et al. Low plasma vitamin B-6 status affects metabolism through the kynurenine pathway in cardiovascular patients with systemic inflammation. *The Journal of nutrition* 2011;141(4):611-7.
273. Platten M, Wick W, Van den Eynde BJ. Tryptophan catabolism in cancer: beyond IDO and tryptophan depletion. *Cancer research* 2012;72(21):5435-40.



274. Sakakeeny L, Roubenoff R, Obin M, Fontes JD, Benjamin EJ, Bujanover Y, et al. Plasma pyridoxal-5-phosphate is inversely associated with systemic markers of inflammation in a population of U.S. adults. *The Journal of nutrition* 2012;142(7):1280-5.
275. Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, et al. Mendelian randomization of blood lipids for coronary heart disease. *European heart journal* 2015;36(9):539-50.
276. Binder AM, Michels KB. The causal effect of red blood cell folate on genome-wide methylation in cord blood: a Mendelian randomization approach. *BMC bioinformatics* 2013;14:353.
277. Chanarin I. *Folate deficiency*. In: Blakley RL, Whitehead VM, eds. *Folates and pterins*. New-York: John Wiley & Sons, 1986.
278. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials* 1998;19(1):61-109.
279. Hays J, Hunt JR, Hubbell FA, Anderson GL, Limacher M, Allen C, et al. The Women's Health Initiative recruitment methods and results. *Ann Epidemiol* 2003;13(9 Suppl):S18-77.
280. Signorello LB, Hargreaves MK, Steinwandel MD, Zheng W, Cai Q, Schlundt DG, et al. Southern community cohort study: establishing a cohort to investigate health disparities. *J Natl Med Assoc* 2005;97(7):972-9.
281. Prorok PC, Andriole GL, Bresalier RS, Buys SS, Chia D, Crawford ED, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials* 2000;21(6 Suppl):273S-309S.
282. Calle EE, Rodriguez C, Jacobs EJ, Almon ML, Chao A, McCullough ML, et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. *Cancer* 2002;94(2):500-11.
283. Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* 2000;151(4):346-57.
284. Buring JE HC. The women's health study: Rationale and background. *J Myocardial Ischemia* 1992;4:30-40.
285. Final Report on the Aspirin Component of the Ongoing Physician's Health Study. *NEJM* 1989 July;321:129-35.
286. Christen WG, Gaziano JM, Hennekens CH. Design of Physicians' Health Study II--a randomized trial of beta-carotene, vitamins E and C, and multivitamins, in prevention of cancer, cardiovascular disease, and eye disease, and review of results of completed trials. *Ann Epidemiol* 2000;10(2):125-34.
287. Belanger CF, Hennekens CH, Rosner B, Speizer FE. The nurses' health study. *The American journal of nursing* 1978;78(6):1039-40.
288. Colditz GA, Hankinson SE. The Nurses' Health Study: lifestyle and health among women. *Nat Rev Cancer* 2005;5(5):388-96.
289. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985;122(1):51-65.
290. Wu K, Willett WC, Fuchs CS, Colditz GA, Giovannucci EL. Calcium intake and risk of colon cancer in women and men. *J Natl Cancer Inst* 2002;94(6):437-46.
291. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *Am J Epidemiol* 1992;135(10):1114-26; discussion 27-36.
292. 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. *International Journal of Epidemiology* 1989;18(4):858-67.
293. 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. *Journal of the American Dietetic Association* 1993;93:790-96.
294. Giovannucci E, Ascherio A, Rimm EB, Colditz GA, Stampfer MJ, Willett WC. Physical activity, obesity, and risk for colon cancer and adenoma in men. *Ann Intern Med* 1995;122(5):327-34.
295. 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(7):790-6.
296. Giles GG, English DR. The Melbourne Collaborative Cohort Study. *IARC Sci Publ* 2002;156:69-70.

297. Berglund G, Elmstahl S, Janzon L, Larsson SA. The Malmo Diet and Cancer Study. Design and feasibility. *J Intern Med* 1993;233(1):45-51.
298. Hallmans G, Agren A, Johansson G, Johansson A, Stegmayr B, Jansson JH, et al. Cardiovascular disease and diabetes in the Northern Sweden Health and Disease Study Cohort - evaluation of risk factors and their interactions. *Scand J Public Health Suppl* 2003;61:18-24.
299. The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. The ATBC Cancer Prevention Study Group. *Ann Epidemiol* 1994;4(1):1-10.
300. Shu XO, Li H, Yang G, Gao J, Cai H, Takata Y, et al. Cohort Profile: The Shanghai Men's Health Study. *International journal of epidemiology* 2015;44(3):810-8.
301. Zheng W, Chow WH, Yang G, Jin F, Rothman N, Blair A, et al. The Shanghai Women's Health Study: rationale, study design, and baseline characteristics. *Am J Epidemiol* 2005;162(11):1123-31.
302. Hankin JH, Stram DO, Arakawa K, Park S, Low SH, Lee HP, et al. Singapore Chinese Health Study: development, validation, and calibration of the quantitative food frequency questionnaire. *Nutr Cancer* 2001;39(2):187-95.
303. Ross RK, Yuan JM, Yu MC, Wogan GN, Qian GS, Tu JT, et al. Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma. *Lancet* 1992;339(8799):943-6.
304. Yuan JM, Ross RK, Wang XL, Gao YT, Henderson BE, Yu MC. Morbidity and mortality in relation to cigarette smoking in Shanghai, China. A prospective male cohort study. *Jama* 1996;275(21):1646-50.
305. Ueland PM, Ulvik A, Rios-Avila L, Midttun O, Gregory JF. Direct and Functional Biomarkers of Vitamin B6 Status. *Annu Rev Nutr* 2015;35:33-70.
306. Bailey LB, Stover PJ, McNulty H, Fenech MF, Gregory JF, 3rd, Mills JL, et al. Biomarkers of Nutrition for Development-Folate Review. *J Nutr* 2015;145(7):1636S-80S.
307. Midttun O, Hustad S, Ueland PM. Quantitative profiling of biomarkers related to B-vitamin status, tryptophan metabolism and inflammation in human plasma by liquid chromatography/tandem mass spectrometry. *Rapid Commun.Mass Spectrom.* 2009;23(9):1371-79.
308. Ueland PM, Midttun O, Windelberg A, Svardal A, Skalevik R, Hustad S. Quantitative profiling of folate and one-carbon metabolism in large-scale epidemiological studies by mass spectrometry. *Clin.Chem.Lab Med.* 2007;45(12):1737-45.
309. Midttun O, Kvalheim G, Ueland PM. High-throughput, low-volume, multianalyte quantification of plasma metabolites related to one-carbon metabolism using HPLC-MS/MS. *Anal Bioanal Chem* 2012.
310. Molloy AM, Scott JM. Microbiological assay for serum, plasma, and red cell folate using cryopreserved, microtiter plate method. *Methods Enzymol.* 1997;281:43-53.

## **APPENDICES**

### **Appendix 1- Supplementary methods of LC3 cohorts**

#### **Study population**

We invited all prospective cohort studies that in 2009 were members in the US National Cancer Institute (NCI) Cohort Consortium to participate in the study. Additional inclusion criteria included the occurrence of at least 200 incident lung cancer cases with baseline questionnaire data and either plasma or serum samples cryopreserved at  $<80^{\circ}\text{C}$  available. Twenty cohorts fulfilled those criteria and accepted to participate, resulting in a combined cohort population of over 2,000,000 participants from North America, Europe, Asia and Australia.

#### **Selection of cases and controls**

Lung cancer cases were defined on the basis of the International Classification of Diseases for Oncology, Second Edition (ICD-O-2), and included all invasive cancers coded as C34.0-C34.9. Altogether, 11,399 incident lung cancer cases with pre-diagnostic blood samples were identified from the participating cohorts. We selected a total 5,545 lung cancer cases, and in order to optimize the statistical power in smoking stratified analyses, never and former smoking cases were oversampled. For each case, one control was randomly chosen from risk-sets consisting of all cohort members alive and free of cancer (except non-melanoma skin cancer) at the time of diagnosis of the index case. Matching criteria were cohort, sex, date of blood collection ( $\pm 1$  month, relaxed to  $\pm 3$  months for sets without available controls), and date of birth ( $\pm 1$  year, relaxed to  $\pm 3$  years), as well as smoking status in 5 categories; never smokers, short and long term quitters among former smokers ( $<10$  years,  $\geq 10$  years since quitting), and light and heavy smokers among current smokers ( $< 15$  years,  $\geq 15$  cigarettes per day). After excluding cases who were not correctly matched on smoking status ( $n=124$  cases), who had insufficient plasma samples ( $n=42$ ), or had a revised date of diagnosis prior to blood draw ( $n=13$ ), 5,364 lung cancer case-control pairs remained eligible for the risk analysis.

## **Brief description of the participating cohorts in the Lung Cancer Cohort Consortium**

### **US cohorts**

#### **The Women's Health Initiative (WHI)**

WHI is a long-term health study of 161,808 post-menopausal women aged 50 to 79 years at 40 clinical centers throughout the U.S. WHI comprises a Clinical Trial (CT) component (68,132 women), and an Observational Study (OS) component (93,676 women), and has included several extension studies. Some detailed descriptions of WHI have been previously presented<sup>278 279</sup>. The CT evaluated two forms of postmenopausal hormone therapy, a low-fat dietary pattern intervention, and calcium and vitamin D supplementation in a randomized, controlled fashion, in a partial factorial design. The hormone therapy component findings led to major reductions in the use of hormone therapy worldwide, and are thought to have led to noteworthy reductions in breast cancer incidence.

In the present study lung cancer cases occurring during the follow-up of WHI cohorts since enrolment (1993-1998) among non-smoking women, were matched 1-1 to corresponding non-smoking lung cancer free controls, for serum and DNA analyte comparisons.

#### **The Southern Community Cohort Study (SCCS)**

The Southern Community Cohort Study (SCCS)<sup>280</sup> is a prospective cohort of African and non-African Americans which during 2002-2009 enrolled approximately 86,000 residents aged 40-79 years across 12 southern states. Recruitment occurred mainly at community health centers, institutions providing basic health services primarily to the medically uninsured, so that the cohort includes many adults of lower income and educational status. Each study participant completed a detailed baseline questionnaire, and nearly 90% provided a biologic specimen (approximately 45% a blood sample and 45% buccal cells). Follow-up of the cohort is conducted by linkage to national mortality registers and to state cancer registries. Included in this study are 240 incident African American lung cancer cases and 240 individually matched African American cohort members without lung cancer at the index date selected by incidence density sampling.

#### **Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO)**

The PLCO study, a randomized trial aimed at evaluating the efficacy of screening in reducing cancer mortality, recruited approximately 155,000 men and women age 55 to 74 years from 1992 to 2001<sup>281</sup>.

Screening for lung cancer among participants in the intervention arm included a chest x-ray at baseline followed by either three annual x-rays (for current or former smokers at enrollment) or two annual x-rays (for never smokers); participants in the control arm received routine health care. Screening-arm participants provided data on sociodemographic factors, smoking behavior, anthropometric characteristics, medical history, and family history of cancer, as well as blood samples annually for the first 6 years of the study (baseline [T0] and T1 through T5). Lung cancers were ascertained through annual questionnaires mailed to the participants, and positive reports were followed up by abstracting medical records or death certificates. Follow-up in the trial as of July 2009 was 96.7%.

We conducted a nested case-control study within the screening arm of the PLCO trial. As of December 31, 2004, 898 lung cancers were diagnosed among the 77,464 participants. Patients were excluded because of missing baseline questionnaire, previous history of any cancer, diagnosis of multiple cancers during follow-up, missing smoking information at baseline, missing consent for utilization of biologic specimens for etiologic studies, or unavailability/insufficient quantity of serum or DNA specimens. Hemolyzed vials were excluded. We included 450 confirmed lung cancer patients and 450 matched controls in this study sampled from the intervention arm.

Controls were individuals free of cancer at the time of a case's lung cancer diagnosis. Controls were individually matched to lung cancer patients on sex, date of birth +/- 1 year with a possible relaxation to 5 years, race, study year of blood draw, date of blood draw +/- 1 month (with a possible relaxation to 3/6 months), time of blood draw (6AM-9AM, 9AM-12PM, other), smoking categories (smoking status at enrollment, never, former, or current smoker; cumulative amount of smoking (0 to 29, 30 to 39, 40 to 49, and 50+ pack-years), with additional matching for time since quitting (< 15 years and ≥ 15 years) for former smokers, cigarettes per day smoked, and number of days in the study.

### **The New York University Women's Health Study (NYUWHS)**

The New York University Women's Health Study (NYUWHS) is a prospective cohort study of women enrolled at a mammography screening center in New York City. From March 1985 through June 1991, 14,274 women between the ages of 34 and 65 were enrolled in the study. Because the original focus of the study was endogenous hormones and breast cancer, women who had taken hormone medications in the 6 months preceding baseline enrolment were not eligible for the study.

At the time of enrolment, data on demographics, anthropometric measures, medical history, reproductive and lifestyle variables were collected through self-administered questionnaires after written informed consent was obtained. Incident lung cancer cases were identified through active

follow-up of the cohort conducted with questionnaires mailed approximately every two to four years and record linkages with state tumor registries in New York, New Jersey, and Florida, as well as record linkage with the National Death Index (NDI). Medical records were obtained to verify reported cancer outcomes. A total of 171 incident lung cancer cases and 171 individually matched controls were included for this analysis.

### **The American Cancer Society Cancer Prevention Study-II (CPS-II) Nutrition Cohort**

The ACS Study-II is a prospective study of cancer incidence and mortality among 86,404 men and 97,786 women. The CPS-II Nutrition Cohort, which is described in detail elsewhere<sup>282</sup>, was initiated in 1992 as a subgroup of CPS-II, a prospective study of cancer mortality involving approximately 1.2 million Americans begun in 1982. Participants in the CPS-II Nutrition Cohort were recruited from CPS-II members who resided in 21 states and were between the ages of 50 and 74 years. At enrollment in 1992/1993, participants completed a self-administered questionnaire that included demographic, medical, dietary, and lifestyle information. Follow-up questionnaires were sent to all living Nutrition Cohort members in 1997, and every two years after this to update exposure information and to ascertain newly diagnosed cancers. Between June 1998 and June 2001, blood samples were collected from a subset of CPS-II Nutrition Cohort participants (21,965 women and 17,411 men).

Incident lung cancer cases were identified through self-report on a follow-up questionnaire, linkage with state cancer registries, or death certificates. Self-reported cancers were verified through medical records. 200 incident lung cancer cases and 200 matched controls from the CPS-II Nutrition Cohort were provided for this analysis.

### **The Campaign Against Cancer and Stroke (CLUE I) and the Campaign Against Cancer and Heart Disease (CLUE II).**

The CLUE studies include two large cohorts of volunteers from Washington County, Maryland that were enrolled in 1974 and 1989, respectively. CLUE I was conducted in Washington County, Maryland, in the fall of 1974. Brief health histories and blood pressures were taken and 15 ml of blood was drawn from 26,147 volunteers (23,951 were residents of Washington County) at the time of enrollment. Linkage of the records from this program to those of a private census in the summer of 1975 indicated that almost a third of the adult population of the county had participated. CLUE II was an outgrowth of CLUE I conducted from May through October in 1989. As in CLUE I, a brief health history was obtained and 20 ml of blood was drawn. A blood sample was collected from 32,894 volunteers at the time of enrollment (25,076 were residents of Washington County). Participants were also given a food frequency

questionnaire to complete at home and were asked to return it with a toenail clipping of the large toe for trace metal assays. Comparisons with published figures from the 1990 Census indicated that approximately 30 percent of adult residents had participated.

### **The Multiethnic Cohort (MEC)**

The MEC includes over 215,000 men and women aged 45-75 years at recruitment from five different racial/ethnic groups (African Americans, Japanese Americans, Native Hawaiians, Latinos and European Americans) in Hawaii and California<sup>283</sup>. The cohort was assembled in 1993-1996 by mailing a self-administered, 26-page questionnaire to obtain extensive information on demographics, medical and reproductive histories, medication use, family history of various cancers, physical activity and diet. Identification of incident cancer cases is by regular linkage with the Hawaii, Los Angeles County and California SEER registries. From 1995 to 2001, blood collection was conducted from incident cases with breast, prostate, or colorectal cancers, as well as a random sample of cohort participants to serve as controls in genomic nested case-control studies (participation rate 72% and 63%, respectively). In addition, from 2001 to 2006, blood was also collected prospectively, without regard for cancer diagnosis, from willing cohort participants. Approximately 67,000 gave a blood sample (participation rate 43%). All incident lung cancer cases diagnosed before 2010 with a pre-diagnostic blood sample were considered for inclusion in this study. Each case was matched to a control based on study site, sex, age, race/ethnicity, smoking status, hours of fasting, and date and time of blood draw.

### **Women's Health Study (WHS)**

The WHS was a randomized trial of low-dose aspirin, vitamin E, and beta-carotene in the primary prevention of cardiovascular disease and cancer beginning in 1992 among 39,876 female US health professionals aged  $\geq 45$  years<sup>284</sup>. Information on major clinical, lifestyle, and dietary factors was collected via self-reports on baseline questionnaires. Women also provided baseline bloods. During more than two decades of follow-up, WHS participants reporting new cases of cancer on annual follow-up questionnaires were confirmed by medical record review by the WHS Endpoints Committee. Reports of cancer were confirmed on the basis of pathology or cytology reports or, rarely, strong clinical and radiologic or laboratory marker evidence when a pathology or cytology review was not conducted. Only confirmed cases of lung cancer were included in the present analyses, which were matched with eligible controls.

### **Physicians' Health Study (PHS)**

The PHS I began in 1982 as a randomized trial of aspirin and beta-carotene for the primary prevention of heart disease and cancer among 22,071 male, Caucasian physicians initially aged 40 to 84 years<sup>285</sup>, followed by the PHS II trial beginning in 1997 to evaluate beta-carotene, vitamin C, vitamin E, and a daily multivitamin on the prevention of cancer, CVD, and other endpoints. The PHS II included 14,641 men, with 7,641 participants from the PHS I plus 7,000 new physicians, for a total of 29,071 PHS participants<sup>286</sup>. A wide range of demographic, clinical, and lifestyle factors were assessed via baseline questionnaires, along with baseline bloods. PHS participants reported major clinical endpoints, including cancer, yearly in a mailed questionnaire and postcards every six months. Self-reported, incident lung cancer cases were confirmed through medical record review by the PHS Endpoints Committee in included in the present analyses.

### **The Nurses' Health Study (NHS)**

The Nurses' Health Study (NHS)<sup>287 288</sup> was established in 1976, when 121,700 married female registered nurses aged 30 to 55 years residing in 11 States in the U.S. completed and returned a self-administered questionnaire. Questionnaires have been mailed to participants in both cohorts every 2 years since baseline to collect updated information on demographics, lifestyle factors, medical history, and disease outcomes. A semi quantitative food frequency questionnaire (FFQ) was administered to obtain information on usual dietary intake over the previous year. The reproducibility and validity of the FFQs have been established<sup>289-293</sup>. The FFQ was first administered in 1980 in the NHS, and were repeated almost every 4 years thereafter. For each food item, the questionnaire specified a common serving size and queried respondents on average intake during the previous year; responses in 9 categories ranged from almost never to 6 or more per day. Most nutritional variables measured by these FFQs have been developed, tested, and refined by our group over the past 30 years (<https://regepi.bwh.harvard.edu/health/>).

The follow-up rate has been greater than 90%. The institutional review board at the Brigham and Women's Hospital approved the study. As approved by the committee, return of the questionnaires was considered to imply informed consent. Cases of lung cancer were self-reported by the participants or identified on their death certificates and were subsequently confirmed by medical records.



### **Health Professionals Follow-up Study (HPFS)**

The Health Professionals Follow-up Study (HPFS)<sup>294</sup> is an ongoing cohort study of 51,529 U.S. male professionals who were aged 40 to 75 years at baseline in 1986. Questionnaires have been mailed to participants in both cohorts every 2 years since baseline to collect updated information on demographics, lifestyle factors, medical history, and disease outcomes. The follow-up rate has been greater than 90%. The institutional review board at the Harvard T.H. Chan School Public Health approved this study. As approved by the committee, return of the questionnaires was considered to imply informed consent. A semiquantitative food frequency questionnaire (FFQ) was administered to obtain information on usual dietary intake over the previous year. The FFQ was first administered in 1986 in the HPFS and was repeated almost every 4 years thereafter. The reproducibility and validity of the FFQ have been established<sup>291 295</sup>. For each food item, the questionnaire specified a common serving size and queried respondents on average intake during the previous year; responses in 9 categories ranged from almost never to 6 or more per day. Cases of lung cancer were self-reported by the participants or identified on their death certificates and were subsequently confirmed by medical records.

#### **1) European/Australian cohorts**

### **The Melbourne Collaborative Cohort Study (MCCS)**

The MCCS is a prospective cohort study of 41,514 participants (17,045 men and 24,469 women) aged 27-88 years at recruitment<sup>296</sup>; 99.3% of whom were aged 40-69 years. Recruitment occurred between 1990 and 1994. Southern European migrants to Australia (including 5,411 Italians and 4,525 Greeks) were over-sampled to extend the range of lifestyle exposures and to increase genetic variation.

Subjects were recruited via Electoral Rolls (registration to vote is compulsory for adults in Australia), advertisements, and community announcements in local media. Comprehensive lists of Italian and Greek surnames were used to target southern European migrants in phonebooks and electoral rolls. Passive follow-up of the cohort has been conducted by record linkage to Electoral Rolls, electronic phonebooks, the Victorian Cancer Registry and death records; as well as national cancer and death records to identify events outside of Victoria.

At recruitment participant's height and weight were measured, blood samples collected and questionnaires covering lifestyle (diet, smoking, physical activity and alcohol consumption), demographics and medical history completed.

Incident lung cancer cases under 80 years of age up to 31<sup>st</sup> December 2010 were identified through record linkage. A total of 361 lung cancer cases and 361 controls matched for date of birth, gender, smoking status, and ethnicity were included for this analysis.

### **The Malmö Diet and Cancer Study (MDCS)**

The Malmö Diet and Cancer Study (MDCS) is a population-based prospective cohort study that between 1991 and 1996 recruited men and women aged 44 to 74 years of age living in Malmö, Sweden<sup>297</sup>. The main goal of the MDCS is to study the impact of diet on cancer incidence and mortality. It consists of a baseline examination including dietary assessment, a self-administered questionnaire, anthropometric measurements and collection of blood samples. A total of 201 incident lung cancer cases and 201 individually matched controls were available for this analysis.

### **The Northern Sweden Health and Disease Study Cohort (NSHDS)**

The Northern Sweden Health and Disease Study (NSHDS) encompasses several prospective cohorts, the current study involving study participants from the Västerbotten Intervention Project (VIP), a sub-cohort within NSHDS<sup>298</sup>. VIP is an ongoing prospective cohort and intervention study intended for health promotion of the general population of the Västerbotten County in northern Sweden. VIP was initiated in 1985 and all residents in the Västerbotten County were invited to participate by attending a health check-up at 40, 50 and 60 years of age. Participants were asked to complete a self-administered questionnaire including various demographic factors such as education, smoking habits, physical activity and diet. In addition, height and weight were measured and participants were asked to donate a fasting blood sample for future research. Incident lung cancer cases were identified through linkage with the regional cancer registry. A total of 245 incident lung cancer cases and 245 individually matched controls were included for this analysis.

### **The Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC)**

The ATBC Study was a randomized, double-blind, placebo-controlled, primary cancer prevention trial testing daily supplementation with  $\alpha$ -tocopherol (50 mg/day) or  $\beta$ -carotene (20 mg/day), or both<sup>299</sup>. Between 1985 and 1988, the study enrolled and randomized 29,133 50-69 year old male cigarette smokers from southwestern Finland. Study supplementation continued for 5-8 years (median 6.1 years) until death or trial closure (April 30, 1993). At baseline, participants completed questionnaires regarding general risk factors, medical history, smoking habits, and dietary intake. Height, weight, heart rate, and blood pressure were measured by trained nurses and fasting serum samples were

collected and stored at  $-70^{\circ}\text{C}$ . Lung cancer cases diagnosed during follow-up through 2009 were identified through linkage with the Finnish Cancer Registry, and 200 were individually matched to 200 controls for this analysis.

### **The Nord-Trøndelag Health Study (HUNT)**

The HUNT study is a longitudinal population based study having invited all persons aged 20-100 years living in the county of Nord-Trøndelag, Norway to three data collections, HUNT1 (1984-86), HUNT2 (1995-97) and HUNT 3 (2006-08) (<http://www.ntnu.edu/hunt>). Comprehensive data on life style, health status, symptoms, diseases and anthropometrics have been collected through questionnaires, interviews and clinical examinations, and in HUNT2 and HUNT3 biological material as blood and urine additionally were collected and stored. For this study, data have been linked to the Norwegian Cancer Registry and 238 incident lung cancer cases and 238 individually matched controls have been included.

## **2) Asian cohorts**

### **The Shanghai Men's Health Study (SMHS) and the Shanghai Women's Health Study (SWHS)**

The SMHS and SWHS are population-based cohort studies conducted in eight communities of urban Shanghai. Their designs and methods have been described elsewhere<sup>300 301</sup>.

Briefly, the SWHS recruited 74,941 women during 1997-2000 (response rate: 93%) and the SMHS recruited 61,480 men during 2002-2006 (response rate: 74%). Similar methods and questionnaires were used in both studies. At baseline in-person interviews, information on sociodemographic, diet, lifestyle, occupation and medical history was obtained; height, body weight, and waist circumference were measured. Blood samples were collected from 75% of the study participants in both studies, processed within 6 hours, and stored at  $-70^{\circ}\text{C}$  until analysis.

The SMHS and SWHS have been followed up by annual record linkage with the population-based Shanghai Cancer Registry and Shanghai Vital Statistics Registry and in-person surveys every 2-3 years. Exposure information, including dietary intake, was updated in the in-person follow-up surveys. All possible matches from the linkages are checked manually and verified by home visits. Medical charts were obtained from the initial diagnostic hospitals to verify cancer diagnosis. Death certificate data from the Shanghai Vital Statistics Unit was used to identify the primary cause of death.

The studies were approved by the Institutional Review boards of the Shanghai Cancer Institute and Vanderbilt University. Informed consent was obtained from all participants.

### **The Singapore Chinese Health Study (SCHS)**

The design of the SCHS study has been described<sup>193 302</sup>. Briefly, the cohort was drawn from permanent residents or citizens of Singapore who resided in government-built housing estates (86% of the Singapore population reside in such facilities). The eligible age range for cohort enrolment was 45-74 years. We restricted study subjects to the two major dialect groups of Chinese in Singapore: the Hokkiens and the Cantonese, who originated from Fujian and Guangdong provinces in Southern China, respectively. Between April 1993 and December 1998, 63,257 subjects (approximately 85% of eligible subjects) were enrolled into the cohort study. At recruitment, each study subject was interviewed in person by a trained interviewer using a structured questionnaire that emphasized current diet assessed via a validated, 165-item food frequency questionnaire. The questionnaire also requested information on demographics, lifetime use of tobacco, incense use, current physical activity, usual sleep duration, reproductive history (women only), occupational exposure, medical history, and family history of cancer.

Beginning in April 1994, a random 3% sample of cohort participants were asked to provide blood or buccal cell (if request for blood sample was denied), and spot urine samples. Eligibility for this biospecimen subcohort was extended to all surviving cohort participants starting in January 2000. By April 2005, all surviving cohort subjects had been contacted for biospecimen donation. Approximately 60% of eligible cohort participants donated biospecimens.

The cohort has been passively followed for death and cancer occurrence through regular record linkage with the population-based Singapore Cancer Registry and the Singapore Registry of Births and Deaths. Migration out of Singapore, especially among housing estate residents, is negligible. As of latest update, only 55 individuals from this cohort were known to be lost to follow-up due to migration and other reason.

A nested case-control study of incident lung cancer cases within the Singapore Chinese Health Study was used to examine the association between serum levels of vitamin B<sub>6</sub> and other compounds in the one-carbon metabolism pathway and risk of lung cancer. Briefly, 422 lung cancer cases were identified among cohort participants with available prediagnostic plasma samples as of 12/31/2011. For each case, one control subject was randomly selected from all eligible cohort members who were alive and free of cancer on the date of cancer diagnosis of the index case. The control subject was individually matched to the index case by gender, dialect group (Hokkien, Cantonese), age at enrolment ( $\pm 3$  years),

date of baseline interview ( $\pm 2$  year), date of biospecimen collection ( $\pm 6$  months), and smoking status (current, former, and never smokers). For current smokers, cases and controls were further matched by number of cigarettes per day ( $< 15$ ,  $\geq 15$  cigarettes/day). For former smokers, cases and controls were further matched by years since quitting smoking ( $< 10$ ,  $\geq 10$  years). One plasma aliquot per subject was retrieved from the biorepository and all plasma samples were sent to the laboratory (B-vital) for measurements.

### **The Shanghai Cohort Study (SCS)**

The SCS study is a residential cohort of 18,244 men in Shanghai, China, assembled during 1986-89 when subjects were between the ages of 45 and 64 years. Approximately 80% of eligible men participated in the study. At the time of recruitment, each cohort subject was interviewed in-person by a trained nurse interviewer using a structured questionnaire that included background information, history of tobacco and alcohol use, current diet, and medical history<sup>303 304</sup>.

At the completion of the interview, the nurse collected a 10 ml blood and a single void urine specimen from the study participant. Blood and urine samples were kept in insulated boxes with ice ( $0-2^{\circ}\text{C}$ ). The serum was separated from blood specimen within 3-4 hours after collection. Two sets of serum (2 ml and 1 ml, respectively) and two sets of urine samples (10 ml each) per subject have been stored at  $-80^{\circ}\text{C}$ .

The cohort has been followed for the occurrence of cancer and death through routine ascertainment of new cases from the population-based Shanghai Cancer Registry and Shanghai Vital Statistics Units. To maximize the cancer findings and minimize the loss of follow-up, we have recontacted each surviving cohort member annually. Retired nurses visit the last known address of each living cohort member and record details of the interim health history of the cohort member. As of December 31, 2014, cumulatively 612 (3.4%) original subjects were lost to follow-up (i.e., persons we have no record of death and we have been unable to locate through our annual follow-up recontacts), and 574 (3.1%) refused to our continued follow-up interview (their cancer and vital status has been continually updated through record linkage analyses) after 26 years of follow-up since the beginning of the study.

A nested case-control study of incident lung cancer cases within the Shanghai Cohort Study was used to examine the association between serum levels of vitamin B<sub>6</sub> and other compounds in the one-carbon metabolism pathway and risk of lung cancer. Briefly, 516 lung cancer cases were identified among

cohort participants with available serum samples as of 12/31/2006. For each case, we randomly selected one control subject from all cohort members who were free of cancer and alive at the time of cancer diagnosis of the index case. Controls were matched to the index case by age at enrolment ( $\pm 2$  years), date of biospecimen collection ( $\pm 1$  month) and neighbourhood of residence at recruitment, and smoking status (current, former and never smokers) as established previously for other studies. For former smokers, cases and controls were further matched by years since quitting smoking ( $<10$  vs  $\geq 10$  years). One serum vial per subject was retrieved from biorepository and all serum samples were sent to the laboratory (B-vital) for measurements.

### Biochemical analyses

Serum or plasma samples from all study participants were sent on dry ice to the Bevital A/S laboratory (<http://www.bevital.no>) in Bergen, Norway, where they were kept below  $-80^{\circ}\text{C}$  until analysis. Concentrations of vitamin B6, methionine and cotinine were determined by mass spectrometry based methods (LC-MS/MS, GC-MS/MS),<sup>239 240</sup> and microbiological methods were used to determine concentrations of folate (*Lactobacillus casei*).<sup>241</sup> Samples were analyzed in batches of 86 samples, and quality control procedures included 6 calibration samples, 2 control samples, and 1 blank sample in each batch. The coefficients of variation (CVs) within and between batches were, respectively, 3% and 7% for vitamin B6, 4% and 5% for folate, and 1% and 3% for methionine. All lung cancer cases and their individually matched controls were analyzed together within the same batches in random order. The laboratory staff was blinded to the case-control status of the blood samples.

### Statistical analyses

We fitted hierarchical linear models to the biomarker concentrations to evaluate between-cohort variability in mean concentrations, and the extent to which this variability could be accounted for by individual level covariates such as age, sex, body mass index, and smoking status. To account for the nested nature of the data (with participants nested within cohorts) and to allow model parameters to vary across cohorts, we used a hierarchical linear regression model to regress standardized, log transformed biomarker concentrations on these covariates. In particular, let  $\mathbf{y}$  denote the outcome variable (i.e., log transformed concentrations), and  $X$  denote a matrix of  $K$  covariates. We model the outcome of participant  $n$  from cohort  $j$  as

$$y_{nj} \sim \text{Normal}(\alpha_j + x_{nj}\boldsymbol{\beta}_j, \sigma^2).$$

Each of the  $K$  components of the cohort-specific vector of regression coefficients  $\beta_{jk}$  are modeled with independent normal distributions, as are the cohort-specific intercepts  $\alpha_j$ :

$$\beta_{jk} \sim \text{Normal}(\mu_k, \tau_k^2); \alpha_j \sim \text{Normal}(\phi, \delta^2).$$

The overall intercept  $\phi$  and mean regression coefficients  $\mu_k$  were given diffuse  $\text{Normal}(0, 50^2)$  prior distributions, and the hierarchical standard deviation parameters  $\tau_k$  and  $\delta$  were given weakly informative  $\text{HalfCauchy}(0, 2)$  hyperprior distributions. The residual standard deviation  $\sigma$  was also given a  $\text{HalfCauchy}(0, 2)$  prior distribution. We conducted sensitivity analyses in which we specified different scales for the Half-Cauchy hyperpriors, and found that the posterior distributions of  $\alpha_j$  and  $\beta_{jk}$  were insensitive to the specified scale of the hyperpriors.

Risk analysis involved calculating quartiles of circulating concentrations for each biomarker based on the distribution among controls, both overall and by region (United States [US], Europe and Australia combined [EU/AU], and Asia). Relative risks of lung cancer were estimated by calculating odds ratios (OR) and 95% confidence intervals (CI) using conditional logistic regression (conditioning on matched case-control sets) with the first quartile as the referent. Additional covariates were included to account for confounding by risk factors, including indicators of educational attainment (in six categories) and tobacco exposure (in addition to matching to smoking status in 5 categories: cotinine concentrations [defined using quartiles of the distribution among participants reporting to be current smokers]). Including additional covariates of body mass index (BMI) and alcohol intake did not appreciably alter the results and were not included in the final models. As a sensitivity analysis, we fitted models that were additionally adjusted for smoking duration or pack-years of smoking among ever smokers, 90% of which had such information available. All risk analyses were conducted overall and by stratifying for never, former and current smokers, as well as by region (US, EU/AU, Asia).

As an indication of the overall statistical strength of association between each biomarker and risk, we calculated a p-value for trend by including the base-2 logarithm ( $\log_2$ ) of the biomarker concentration as a continuous variable in a separate conditional logistic regression model. The same approach was used in stratified risk analyses according to other pre-defined demographic characteristics and risk factors. OR estimates per  $\log_2$  unit ( $\log_2$  OR) may be interpreted as the relative risk associated with a doubling in the concentration of a circulating biomarker. We also performed risk analyses for  $\log_2$ OR on each biomarker separately for each individual cohort, as well for US cohorts by the period of blood sample collection (before and after 1996 when folate fortification of food items took effect) to evaluate

potential impact of folate fortification.  $\chi^2$  tests were used to assess heterogeneity in  $\log_2$  OR estimates in stratified analyses.

Finally, in order to evaluate the specific hypothesis of whether deficiency in vitamin B6 or folate is particularly important in lung cancer etiology, we conducted analysis by categorizing subjects into deficient, or normal category based on clinically defined cut-off values. For vitamin B6,  $<20$  nmol/L was considered deficient whereas  $\geq 20$  nmol/L was considered normal<sup>305</sup>. For folate deficiency, subjects were classified in three categories;  $<7$  nmol/L as deficient, 7-12.9 nmol/L as moderately deficient, and  $\geq 13$  nmol/L as normal<sup>244 306</sup>.

All statistical analyses were conducted using SAS 9.2 (Cary, North Carolina)<sup>245</sup>, R version 3.1.3<sup>247</sup> or Stan version 2.9.0<sup>248</sup>.



Appendix 2-Data dictionary of LC3 variables

**Reference data specification sheet for the Lung Cancer Cohort Consortium (LC3)**

**GENERAL INSTRUCTIONS**

This Data Specifications sheet outlines the covariates and formatting of the Lung Cancer Cohort Consortium (LC3) study database. The purpose of this document is to allow each cohort to provide harmonized covariates. The dictionary was adapted from that used in the Vitamin D Pooling Project (courtesy of Stephanie Smith-Warner and Stephanie Weinstein and the VDPP). The centralized LC3 database will be in SAS format (\*.sas7bdat) and managed by the International Agency for Research on Cancer (IARC).

To the furthest extent possible, please format your study database as described below and send it to Aurelie MOSKAL (moskala@iarc.fr) and Mattias Johansson (johanssonm@iarc.fr). Please include a modified version of this excel file and indicate any changes to data formats required for your specific cohort. In particular, for any variable that cannot be created based on the available data of your cohort database, put "N/A" in the "unit" column. Furthermore, if you have variables that clearly correspond to those below but cannot be formatted as requested, please include such variables and modify the "Format and/or List of Valid Values" accordingly.

Should you prefer uploading the data on e.g. our FTP server contact Aurelie or instructions, as well as for any queries that you may have.

We much appreciate your help in preparing the LC3 study database.

**FILE NAMING CONVENTION AND FILE TYPE**

File Naming Convention (without the brackets): "[cohort].LC3.data.[yyyymmdd].ext" (Ex. "ATBC.LC3.data.20120714.sas7bdat" or "ATBC.LC3.data.20120714.csv")

File type: A comma-delimiter text file or a SAS file is requested.

**TIMING OF DATA COLLECTION**

If not all variables were collected at time of blood draw, it is important to record the timing of data collection. We have included variables with the same root name and "\_YR" to indicate year of collection. As given below they refer to an entire group of similar variables. If you need to specify different time points for individual variables, you can create your own "\_YR" variables using the same root as the variable you are recording. Please see below for definitions of time points:

**any** - Data can be collected at any time.

**blood draw** - Data were collected at time of blood draw.

**closest to blood draw** - Preference is for data collected at blood draw. If not available, then choose data from a time point as close as possible to blood draw. Data could be collected after blood draw as long as the collection was prior to case diagnosis.

**most recent** - The most recent data collection where these data are available.

**index date or most recent prior to index date** - Date of diagnosis for the case, date of case diagnosis for the matched control. Only requested for specific variables, e.g. smoking variables.

**MISSING VALUE CODES**

. = missing/unknown value [the question was asked, but data are missing/unknown for subject(s)]

.U = unavailable [the question was not asked]

.N = not applicable [the question does not apply to the subject]

**NOTE:** The . and .U codes may be used in any variable. The .N code may be used when provided in the list of valid values.

**Basic cohort information**

BASELINE_EXCL_CRITERIA	Description of baseline exclusion criteria (as determined by date of blood draw, common for all subjects per cohort). For example, indicate whether participants were excluded from the data set at baseline if they had a history of any cancer, history of lung cancer only	any	free text	
FOLLOW-UP_EXCL_CRITERIA	Description of whether individuals were excluded as cases and potential controls if they were diagnosed with a different cancer after baseline and during follow-up prior to diagnosis of the cancer of interest	any	free text	

MATCH_DESCRIPTION	Description of variables used as matching factors. This would be particularly important if the matching protocol differed from the LC3 study protocol			Free text including the variable names and actual categories used for the matching factors. If those variables are not already included in the variable list, please add the matching variables and their corresponding formats to the data set  If the matching factors for your study are not included in the variable list, please include those variables, a description of those variables, and the formatting for valid values for those variables in the data set.	
VARIABLE LIST					
Variable Name	Variable Description	Data Collection Time	Format and/or List of Valid Values	Unit	
BASELINE INFORMATION					
GENERAL INFORMATION					
COHORT_ACCR	Cohort acronym	NA	Free text	NA	
LC3_ID	Participant ID (unique participant identifier) using sample ID Format "cohortnumber_serialIDnumber" (EX for ATBC: LC3_ID is 16_0001), the LC3_ID should correspond to the ID variable of the cohort specific covariate database	any	cohortnumber_serial ID number (EX: 16_0001)		
COHORT_ID	Cohort internal ID variable	any	free cohort number		
CASESET	Caseset ID for the matched case-control pair	any	cohortnumber_serial ID number (EX: 16_0001)		
GENDER	Gender	any	1 = male2 = female		
DOB	Date of birth	any	YYYYMMDD		
RACE	Race	any	1 = white 2 = black 3 = asian 4 = other		
RACE_ADDITION	Description of other race	any	free text (e.g. "Pacific Islander", hispanic, african, american)		
STUDYCTR	Study center	any	defined by individual studies		
DRAWDATE	Date of blood draw	blood draw	YYYYMMDD		
DRAWAGE	Age at blood draw	blood draw	integer		
DATE_RECR	Date of recruitment	any	YYYYMMDD		
AGE_RECR	Age at recruitment	any	integer		

DRAWTIME	Time of blood draw	blood draw	HHMM using 24-hour time
FASTING	Subject was fasting prior to blood draw (for at least 8 hours)	blood draw	0 = no 1 = yes
FASTTIME	Amount of time spent fasting prior to blood draw (hours)	blood draw	continuous numeric hours
DRAWTOCENT	Amount of time between blood draw and centrifuge/processing (minutes)	blood draw	continuous numeric minutes
BIRTHPLACE	Birthplace	any	two-letter US state abbreviation if in USA; full name of country if not in USA (e.g. "VA", "MD", "CA", "CHINA", "FINLAND")
<b>FAMILY HISTORY (MOST RECENT PRIOR TO INDEX DATE)</b>			
FAMHX_ANYCA_MR	Family history of any cancer (first degree relatives), most recent	index date or most recent prior to index date	0 = no 1 = yes
FAMHX_ANYCA_NUMBER	Number of first degree relative with other cancer, most recent (if FAMHX_ANYCA_MR=Yes)	index date or most recent prior to index date	Continuous numeric
FAMHX_LUNG	Family history of lung cancer (first degree relatives)	index date or most recent prior to index date	0 = no 1 = yes
FAMHX_LUNG_NUMBER	Number of first degree relative with lung cancer (if FAMHX_LUNG=Yes)	index date or most recent prior to index date	Continuous numeric
FAMHX_YR	Year of data collection (FAMILY HISTORY data)	index date or most recent prior to index date	YYYY
<b>MEDICATION USE (AT TIME OF BLOOD DRAW)</b>			
MEDS_NSAIDS	Current use of aspirin/NSAIDS	closest to blood draw	0 = no 1 = yes
MEDS_NSAIDS_DUR	Duration of current use of aspirin/NSAIDS (years)	closest to blood draw	integer if available; if collected as categorical, report the range as free text .N = not a current user
MEDS_HIGHBP	Current use of high blood pressure medication	closest to blood draw	0 = no 1 = yes
MEDS_DIAB	Current use of diabetes medication	closest to blood draw	0 = no 1 = yes
MEDS_HBURN	Current use of heartburn medication	closest to blood draw	0 = no 1 = yes
MEDS_YR	Year of data collection (MEDS data)	closest to blood draw	YYYY
<b>HISTORY OF DISEASE (AT TIME OF BLOOD DRAW)</b>			
HX_ANYCANCER	History of any cancer (except nonmelanoma skin cancer)	closest to blood draw	0= no 1 =yes
HX_LUNGCANCER	History of Lung cancer	closest to blood draw	0= no 1 =yes

HX_DIAB	History of adult onset diabetes	closest to blood draw	0 = no 1 = yes
HX_HIGHBP	History of high blood pressure	closest to blood draw	0 = no 1 = yes
HX_YR	Year of data collection (HX data)	closest to blood draw	YYYY
<b>LIFESTYLE INFORMATION</b>			
<b>GENERAL INFORMATION</b>			
EDUCATION	Highest Education level	closest to blood draw	1 = less than high school 2 = completed high school 3 = vocational school 4 = some college 5 = college graduate 6 = graduate studies
EDUCATION_YR	Year of data collection (EDUCATION data)	closest to blood draw	YYYY
INCOME_VARIABLE	Information on income to be included when available (Each type allowed), please send the information you have with a description of the items and the units. We will then set up common variables. If several variables are available, please include all of them.	closest to blood draw	free text
INCOME_YR	Year of data collection (INCOME data)	closest to blood draw	YYYY
RESIDENCE	Geographic region where subject resides	closest to blood draw	two-letter US state abbreviation if in USA; full name of country if not in USA (e.g. "VA", "MD", "CA", "ITALY", "FINLAND")
RESIDENCE_YR	Year of data collection (RESIDENCE data)	closest to blood draw	YYYY
<b>SMOKING (AT TIME OF BLOOD DRAW)</b>			
SMOKE_STATUS	Smoking status (all types of tobacco)	closest to blood draw	If collected as never/ever: 0 = never 1 = ever If collected as current (no/yes): 2 = no current 3 = current If collected as never/former/current: 4 = never 5 = former 6 = current

SMOKE_STATUS_CIGARETTE	Cigarette smoking status	closest to blood draw	<p>If collected as never/ever:                      0 = never                      1 = ever                      If collected as current (no/yes):                      2 = no current                      3 = current                      If collected as never/former/current:                      4 = never                      5 = former                      6 = current</p>	
SMOKE_INT_CIGARETTE	Smoking intensity (number of cigarettes per day CPD) (note: we would like these data for individuals who indicated they were ever smokers, current smokers, or former smokers)	closest to blood draw	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_DUR_CIGARETTE	Smoking duration (number of years smoked cigarettes regularly)	closest to blood draw	continuous numeric if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_PACKYRS_CIGARETTE	Cigarette smoking pack years	closest to blood draw	Continuous numeric .N = subject was never a smoker	
SMOKE_AGE_START_CIGARETTE	Age at which the subject started cigarette smoking	closest to blood draw	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_AGE_QUIT_CIGARETTE	Age at which the subject stopped cigarette smoking	closest to blood draw	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker or did not quit	
SMOKE_CIGAR	Cigar smoking status	closest to blood draw	0 = never 1 = ever	
SMOKE_INT_CIGAR	Smoking intensity (number of cigars per day) (note: we would like these data for individuals who indicated they were ever smokers, current smokers, or former smokers)	closest to blood draw	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_DUR_CIGAR	Smoking duration (number of years smoked cigars regularly)	closest to blood draw	continuous numeric if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_PACKYRS_CIGAR	Cigar smoking pack years	closest to blood draw	Continuous numeric .N = subject was never a smoker	
SMOKE_AGE_START_CIGAR	Age at which the subject started cigar smoking	closest to blood draw	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_AGE_QUIT_CIGAR	Age at which the subject stopped cigar smoking	closest to blood draw	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker or did not quit	

SMOKE_PIPE	Pipe smoking status	closest to blood draw	0 = never 1 = ever	
SMOKE_INT_PIPE	Smoking intensity (number of pipes per day CPD) (note: we would like these data for individuals who indicated they were ever smokers, current smokers, or former smokers)	closest to blood draw	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_DUR_PIPE	Smoking duration (number of years smoked pipes regularly)	closest to blood draw	continuous numeric if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_PACKYRS_PIPE	Pipe smoking pack years	closest to blood draw	Continuous numeric .N = subject was never a smoker	
SMOKE_AGE_START_PIPE	Age at which the subject started pipe smoking	closest to blood draw	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_AGE_QUIT_PIPE	Age at which the subject stopped pipe smoking	closest to blood draw	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker or did not quit	
SMOKE_YR	Year of data collection (SMOKE data)	closest to blood draw	YYYY	
<b>SMOKING (MOST RECENT PRIOR TO INDEX DATE)</b>				
SMOKE_STATUS_MR	Smoking status (all types of tobacco)	index date or most recent prior to index date	If collected as never/ever: 0 = never 1 = ever If collected as current (no/yes): 2 = no current 3 = current If collected as never/former/current: 4 = never 5 = former 6 = current	
SMOKE_STATUS_CIGARETTE_MR	Cigarette smoking status	index date or most recent prior to index date	If collected as never/ever: 0 = never 1 = ever If collected as current (no/yes): 2 = no current 3 = current If collected as never/former/current: 4 = never 5 = former 6 = current	

SMOKE_INT_CIGARETTE_MR	Smoking intensity (number of cigarettes per day CPD) (note: we would like these data for individuals who indicated they were ever smokers, current smokers, or former smokers)	index date or most recent prior to index date	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_DUR_CIGARETTE_MR	Smoking duration (number of years smoked cigarettes regularly)	index date or most recent prior to index date	continuous numeric if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_PACKYRS_CIGARETTE_MR	Cigarette smoking pack years	index date or most recent prior to index date	Continuous numeric .N = subject was never a smoker	
SMOKE_AGE_START_CIGARETTE_MR	Age at which the subject started cigarette smoking	index date or most recent prior to index date	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_AGE_QUIT_CIGARETTE_MR	Age at which the subject stopped cigarette smoking	index date or most recent prior to index date	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker or did not quit	
SMOKE_CIGAR_MR	Cigar smoking status	index date or most recent prior to index date	0 = never 1 = ever	
SMOKE_INT_CIGAR_MR	Smoking intensity (number of cigars per day) (note: we would like these data for individuals who indicated they were ever smokers, current smokers, or former smokers)	index date or most recent prior to index date	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_DUR_CIGAR_MR	Smoking duration (number of years smoked cigars regularly)	index date or most recent prior to index date	continuous numeric if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_PACKYRS_CIGAR_MR	Cigar smoking pack years	index date or most recent prior to index date	Continuous numeric .N = subject was never a smoker	
SMOKE_AGE_START_CIGAR_MR	Age at which the subject started cigar smoking	index date or most recent prior to index date	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_AGE_QUIT_CIGAR_MR	Age at which the subject stopped cigar smoking	index date or most recent prior to index date	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker or did not quit	
SMOKE_PIPE_MR	Pipe smoking status	index date or most recent prior to index date	0 = never 1 = ever	
SMOKE_INT_PIPE_MR	Smoking intensity (number of pipes per day CPD) (note: we would like these data for individuals who indicated they were ever smokers, current smokers, or former smokers)	index date or most recent prior to index date	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker	
SMOKE_DUR_PIPE_MR	Smoking duration (number of years smoked pipes regularly)	index date or most recent prior to index date	continuous numeric if available; if collected as categorical, report the range as free text .N = subject was never a smoker	



SMOKE_PACKYRS_PIPE_MR	Pipe smoking pack years	index date or most recent prior to index date	Continuous numeric .N = subject was never a smoker
SMOKE_AGE_START_PIPE_MR	Age at which the subject started pipe smoking	index date or most recent prior to index date	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker
SMOKE_AGE_QUIT_PIPE_MR	Age at which the subject stopped pipe smoking	index date or most recent prior to index date	integer if available; if collected as categorical, report the range as free text .N = subject was never a smoker or did not quit
SMOKE_YR_MR	Year of data collection (SMOKE_MR data)	index date or most recent prior to index date	YYYY
<b>PHYSICAL ACTIVITY (AT TIME OF BLOOD DRAW)</b>			
PHYSACT_METS_LEISURE	Current leisure time activity, MET hrs/week	closest to blood draw	continuous numeric hrs/week
PHYSACT_HRS_LEISURE	Current leisure time activity, avg. hrs/week	closest to blood draw	continuous numeric hrs/week
PHYSACT_INT_LEISURE	Intensity of current leisure time activity (if METs not available)	closest to blood draw	1 = light 2 = moderate 3 = vigorous .N = METs data are provided
PHYSACT_METS_OCCUP	Current occupational activity, MET hrs/week	closest to blood draw	continuous numeric hrs/week
PHYSACT_HRS_OCCUP	Current occupational activity, avg. hrs/week	closest to blood draw	continuous numeric hrs/week
PHYSACT_INT_OCCUP	Intensity of current occupational activity (if METs not available)	closest to blood draw	1 = light 2 = moderate 3 = vigorous .N = METs data are provided
PHYSACT_METS_TOTAL	Current total activity, MET hrs/week (if leisure/occupation is not specified)	closest to blood draw	continuous numeric
PHYSACT_HRS_TOTAL	Current total activity, avg. hrs/week (if leisure/occupation is not specified)	closest to blood draw	continuous numeric
PHYSACT_INT_TOTAL	Intensity of current total activity (if METs not available; if leisure/occupation is not specified)	closest to blood draw	1 = light 2 = moderate 3 = vigorous .N = METs data are provided
PHYSACT_YR	Year of data collection (PHYSACT data)	closest to blood draw	YYYY
<b>SUN EXPOSURE/SUSCEPTIBILITY</b>			
EYECOLOR	Eye color	any	1 = brown or black 2 = blue 3 = grey 4 = green 5 = other

HAIRCOLOR	Hair color	any	1 = black 2 = blonde 3 = light brown 4 = medium or dark brown 5 = red	
SUNBURN	Blistering sun burn	closest to blood draw	0 = never 1 = ever	
SUNBURN_YR	Year of data collection (SUNBURN)	closest to blood draw	YYYY	
<b>ANTHROPOMETRIC MEASURES</b>				
HEIGHT	Height (cm)	any	continuous numeric	cm
HEIGHT_SOURCE	Source of height data	any	1 = self-reported 2 = measured	
HEIGHT_YR	Year of data collection (Height data)	closest to blood draw	YYYY	
WEIGHT	Weight (kg)	closest to blood draw	continuous numeric	kg
WEIGHT_SOURCE	Source of weight data	closest to blood draw	1 = self-reported 2 = measured	
WEIGHT_YR	Year of data collection (WEIGHT data)	closest to blood draw	YYYY	
WHR	Waist-to-hip ratio (WHR)	closest to blood draw	continuous numeric	
WHR_SOURCE	Source of WHR data	closest to blood draw	1 = self-reported 2 = measured	
WHR_YR	Year of data collection (WHR data)	closest to blood draw	YYYY	
WEIGHT_YA	Weight (kg) in young adulthood	any	continuous numeric	kg
WEIGHT_YA_AGE	Age for which weight in young adulthood was assessed (eg, 18, 20)	any	continuous numeric	
WEIGHT_YA_YR	Year of data collection (WEIGHT_YA data)	any	YYYY	
<b>DIETARY INFORMATION</b>				
<b>FOOD AND NUTRIENT INTAKE (AT TIME OF BLOOD DRAW)</b>				
DIET_KCAL	Current diet: total caloric/energy intake including alcoholic beverages, kcal/day	closest to blood draw	continuous numeric	kcal/day
TYPE_DIETQUEST	Type of dietary Questionnaire (EX: FFQ, semi-quantitative dietary questionnaire, 24h-recalls, 7-day records, food propensity questionnaire)	closest to blood draw	free text	
DIET_PROT	Current diet: total protein, g/day (average)	closest to blood draw	continuous numeric	g/day
DIET_PROTAN	Current diet: protein from animal sources, g/day	closest to blood draw	continuous numeric	g/day
DIET_PROTPL	Current diet: protein from plant sources, g/day	closest to blood draw	continuous numeric	g/day

DIET_CARB	Current diet: carbohydrates, g/day (average)	closest to blood draw	continuous numeric	g/day
DIET_FAT	Current diet: total fat, g/day (average)	closest to blood draw	continuous numeric	g/day
DIET_SFAT	Current diet: saturated fat, g/day (average)	closest to blood draw	continuous numeric	g/day
DIET_MFAT	Current diet: monounsaturated fat, g/day (average)	closest to blood draw	continuous numeric	g/day
DIET_PFAT	Current diet: polyunsaturated fat g/day (average)	closest to blood draw	continuous numeric	g/day
DIET_CHORL	Current diet: Cholesterol g/day	closest to blood draw	continuous numeric	g/day
DIET_AFAT	Current diet: animal fat, g/day (includes fat from all animal sources including meat, dairy products and eggs)	closest to blood draw	continuous numeric	g/day
DIET_PLFAT	Current diet: fat plant sources, g/day	closest to blood draw	continuous numeric	g/day
DIET_STARCH	Current diet: starch g/day	closest to blood draw	continuous numeric	g/day
DIET_SUGAR	Current diet: sugars g/day	closest to blood draw	continuous numeric	g/day
DIET_DIETARY_FIBER	Current diet: dietary fiber g/day	closest to blood draw	continuous numeric	g/day
DIET_VITD	Current diet: vitamin D from foods, µg/day	closest to blood draw	continuous numeric	µg/day
DIET_CA	Current diet: calcium from foods, mg/day	closest to blood draw	continuous numeric	mg/day
DIET_FOL	Current diet: folate from foods, µg/day	closest to blood draw	continuous numeric	µg/day
DIET_TOTAL_FOL	Current diet: folate from foods & supplements, µg folate equivalents/day	closest to blood draw	continuous numeric	µg/day
DIET_METHIONINE	Current diet: methionine from foods, mg/day	closest to blood draw	continuous numeric	mg/day
DIET_TOTAL_METHIONINE	Current diet: methionine from foods & supplements, mg/day	closest to blood draw	continuous numeric	mg/day
DIET_MG	Current diet: magnesium from foods, mg/day	closest to blood draw	continuous numeric	mg/day
DIET_P	Current diet: phosphorus from foods, mg/day	closest to blood draw	continuous numeric	mg/day
DIET_RET	Current diet: retinol from foods, µg/day	closest to blood draw	continuous numeric	µg/day
DIET_BETACAROTENE	Current diet: beta-carotene from foods, µg/day	closest to blood draw	continuous numeric	µg/day
DIET_IRON	Current diet: iron from foods, mg/day	closest to blood draw	continuous numeric	mg/day
DIET_POTASSIUM	Current diet: potassium from foods, mg/day	closest to blood draw	continuous numeric	mg/day
DIET_RIBOFLAVIN	Current diet: riboflavin (B2) from foods, mg/day	closest to blood draw	continuous numeric	mg/day
DIET_THIAMIN	Current diet: thiamin (B1) from foods, mg/day	closest to blood draw	continuous numeric	mg/day
DIET_VITB6	Current diet: vitamin (B6) from foods, mg/day	closest to blood draw	continuous numeric	mg/day
DIET_VITB12	Current diet: vitamin (B12) from foods, µg/day	closest to blood draw	continuous numeric	µg/day
DIET_VITC	Current diet: vitamin C from foods, mg/day	closest to blood draw	continuous numeric	mg/day
DIET_VITE	Current diet: vitamin E from foods, mg/day	closest to blood draw	continuous numeric	mg/day
FOOD_DAIRY	Current diet: total dairy (includes milk, cheese, yogurt, butter, margarine, creamers, ice cream, frozen yogurt, chocolate milk, and milkshakes), g/day (average)	closest to blood draw	continuous numeric	g/day

FOOD_VEG	Current diet: vegetables (EXCLUDING POTATOES), g/day	closest to blood draw	continuous numeric	g/day
FOOD_LEAFYVEG	Current diet: leafy vegetables (lettuce, spinach, Swiss chard leaf, watercress), g/day	closest to blood draw	continuous numeric	g/day
FOOD_CRUCIF	Current diet: total cruciferous vegetables (broccoli, cauliflower, kale, cabbage, etc), g/day (average)	closest to blood draw	continuous numeric	g/day
FOOD_OLIVES	Current diet: olives intake (green olives, black olives, stuffed olive), g/day	closest to blood draw	continuous numeric	g/day
FOOD_FRUIT_NUTS	individual intake of fruits, nuts and seeds, g/day	closest to blood draw	continuous numeric	g/day
FOOD_FRUITS	individual intake of fruits, g/day	closest to blood draw	continuous numeric	g/day
FOOD_CEREALS	Current diet: cereal and cereal products intake, g/day	closest to blood draw	continuous numeric	g/day
FOOD_FORTCER	Current diet: fortified cereals (Total, Product 19, Most, etc.), g/day (average)	closest to blood draw	continuous numeric	g/day
FOOD_MEAT	individual intake of meat and meat products (including processed meat), g/day	closest to blood draw	continuous numeric	g/day
FOOD_REDMEAT	individual intake of red meat, g/day	closest to blood draw	continuous numeric	g/day
FOOD_POULTRY	individual intake of poultry, g/day	closest to blood draw	continuous numeric	g/day
FOOD_PROCMEAT	Current diet: total processed meat (lunchmeat, bacon, sausage, etc.), g/day (average)	closest to blood draw	continuous numeric	g/day
FOOD_FISH_SHELLFISH	individual intake of fish and shellfish, g/day	closest to blood draw	continuous numeric	g/day
FOOD_FISH	individual intake of fish, g/day	closest to blood draw	continuous numeric	g/day
FOOD_OFISH	Current diet: fatty/oily fish (dark meat fish such as salmon, tuna, mackerel, etc.), g/day (average)	closest to blood draw	continuous numeric	g/day
FOOD_SOY	Current diet: soy, soy milk, tofu, g/day (average)	closest to blood draw	continuous numeric	g/day
DIET_YR	Year of data collection (DIET data)	closest to blood draw	YYYY	
<b>SUPPLEMENT USE (AT TIME OF BLOOD DRAW)</b>				
SUPP_ANY	Any supplement use	closest to blood draw	If collected as never/ever: 0 = never 1 = ever If collected as current use (no/yes): 2 = no current use 3 = current use	

SUPP_MULT	Multivitamin supplement use	closest to blood draw	<p>If collected as never/ever:                      0 = never                      1 = ever                      If collected as current use (no/yes):                      2 = no current use                      3 = current use</p>	
SUPP_VITD	Vitamin D supplement use	closest to blood draw	<p>If collected as never/ever:                      0 = never                      1 = ever                      If collected as current use (no/yes):                      2 = no current use                      3 = current use</p>	
SUPP_VITD_DOSE	Current vitamin D supplement dosage, including contribution from multivitamins (µg/day)	closest to blood draw	continuous numeric .N = current use status unknown	µg/day
SUPP_CA	Calcium supplement use	closest to blood draw	<p>If collected as never/ever:                      0 = never                      1 = ever                      If collected as current use (no/yes):                      2 = no current use                      3 = current use</p>	
SUPP_CA_DOSE	Current calcium supplement dosage, including contribution from multivitamins (mg/day)	closest to blood draw	continuous numeric .N = current use status unknown	mg/day
SUPP_FOL	Folic acid supplement use	closest to blood draw	<p>If collected as never/ever:                      0 = never                      1 = ever                      If collected as current use (no/yes):                      2 = no current use                      3 = current use</p>	
SUPP_FOL_DOSE	Current folic acid supplement dosage, including contribution from multivitamins (µg/day)	closest to blood draw	continuous numeric .N = current use status unknown	µg/day
SUPP_MG	Magnesium supplement use	closest to blood draw	<p>If collected as never/ever:                      0 = never                      1 = ever                      If collected as current use (no/yes):                      2 = no current use                      3 = current use</p>	
SUPP_MG_DOSE	Current magnesium supplement dosage, including contribution from multivitamins (mg/day)	closest to blood draw	continuous numeric .N = current use status unknown	mg/day

SUPP_P	Phosphorous supplement use	closest to blood draw	If collected as never/ever: 0 = never 1 = ever If collected as current use (no/yes): 2 = no current use 3 = current use	
SUPP_P_DOSE	Current phosphorous supplement dosage, including contribution from multivitamins (mg/day)	closest to blood draw	continuous numeric .N = current use status unknown	mg/day
SUPP_METH	Methionine supplement use	closest to blood draw	If collected as never/ever: 0 = never 1 = ever If collected as current use (no/yes): 2 = no current use 3 = current use	
SUPP_METH_DOSE	Current methionine supplement dosage, including contribution from multivitamins	closest to blood draw	continuous numeric .N = current use status unknown	mg/day
SUPP_YR	Year of data collection (SUPP data)	closest to blood draw	YYYY	
<b>ALCOHOL USE (AT TIME OF BLOOD DRAW)</b>				
ALC	Alcohol use (any type)	closest to blood draw	If collected as never/ever: 0 = never 1 = ever If collected as current use (no/yes): 2 = no current use 3 = current use If collected as never/former/current: 4 = never 5 = former 6 = current	
ALC_AGE_START	Age when start drinking alcohol	closest to blood draw	integer	
ALC_AGE_STOP	Age when stop drinking alcohol, if former drinker	closest to blood draw	integer	
ALC_BEER	Current alcohol use (beer)	closest to blood draw	0 = no 1 = yes	
ALC_WINE	Current alcohol use (wine)	closest to blood draw	0 = no 1 = yes	
ALC_LIQ	Current alcohol use (liquor)	closest to blood draw	0 = no 1 = yes	
ALC_SPIRIT	Current alcohol use (spirits)	closest to blood draw	0 = no 1 = yes	

ALC_ETOH	Current alcohol use (any type), ethanol g/day	closest to blood draw	continuous numeric .N = current use status unknown	ethanol g/day
ALC_BEER_ETOH	Current alcohol use (beer), ethanol g/day	closest to blood draw	continuous numeric .N = current use status unknown	ethanol g/day
ALC_WINE_ETOH	Current alcohol use (wine), ethanol g/day	closest to blood draw	continuous numeric .N = current use status unknown	ethanol g/day
ALC_SPIRITS_ETOH	Current alcohol use (spirits), ethanol g/day	closest to blood draw	continuous numeric .N = current use status unknown	ethanol g/day
ALC_LIQ_ETOH	Current alcohol use (liquor), ethanol g/day	closest to blood draw	continuous numeric .N = current use status unknown	ethanol g/day
ALC_YR	Year of data collection (ALC data)	closest to blood draw	YYYY	
<b>DEATH AND FOLLOW-UP INFORMATION</b>				
DEATH	Deceased	most recent	0 = no 1 = yes	
DEATH_DATE	Death date	most recent	MM/DD/YYYY (include slashes); if exact day and/or month are not known, then replace day and/or month with "99". For example, "MM/99/YYYY" or "99/99/YYYY". .N = subject is not deceased	
DEATH_CAUSE	Cause of death	most recent	free text, provided in whatever format is available .N = subject is not deceased	
DEATH_CODING	Coding system used for DEATH_CAUSE	most recent	free text (e.g. "ICD9") .N = subject is not deceased	
DEATH_CODING_SOURCE	Source of the death information	most recent	free text	
DEATH_YR	Year of data collection (DEATH data)	most recent	YYYY	
LAST_FUP	Date of last follow-up	most recent	MM/DD/YYYY (include slashes); if exact day and/or month are not known, then replace day and/or month with "99". For example, "MM/99/YYYY" or "99/99/YYYY". .N = subject is not deceased	
<b>LUNG CANCER OUTCOME INFORMATION</b>				
DIAGNOSISAGE	Age at diagnosis (for cases or of matched case for controls), if applied	index date or most recent prior to index date	continuous numeric	
LUNG_CANCER_CASE	LUNG cancer case	index date	0 = no 1 = yes	
LUNG_CANCER_ICD	LUNG cancer ICD code	index date	ICD-9 or ICD-10 codes, or whatever format is available. .N = subject is a control	
LUNG_CANCER_ICD_CODING	Coding system used for LUNG_CANCER_ICD	index date	Free text	

LUNG_CANCER_DATE	Date of diagnosis of lung cancer	index date	MM/DD/YYYY (include slashes); if exact day and/or month are not known, then replace day and/or month with "99". For example, "MM/99/YYYY" or "99/99/YYYY". .N = subject is a control
LUNG_CANCER_HIST	LUNG cancer histology	index date	WHO ICD-O morphology codes, or whatever format is available. .N = subject is a control
LUNG_CANCER_HIST_CODING	Coding system used for LUNG_CANCER_HIST	index date	free text (e.g. "WHO ICD-O") .N = subject is a control
LUNG_CANCER_GRADE	LUNG cancer grade	index date	whatever format is available .N = subject is a control
LUNG_CANCER_GRADE_CODING	Coding system used for LUNG_CANCER_GRADE	index date	free text (e.g. "AJCC fourth edition") .N = subject is a control
LUNG_CANCER_STAGE	LUNG cancer stage	index date	TNM Stage, or whatever format is available. .N = subject is a control
LUNG_CANCER_STAGE_CODING	Coding system used for LUNG_CANCER_STAGE	index date	free text (e.g. "AJCC fourth edition") .N = subject is a control
<b>ADDITIONAL VARIABLES</b>			
<b>HORMONE THERAPY USE (AT TIME OF BLOOD DRAW)</b>			
HRT	Hormone therapy use	closest to blood draw	If collected as never/ever: 0 = never 1 = ever If collected as current use (no/yes): 2 = no current use 3 = current use If collected as never/former/current: 4 = never 5 = former 6 = current .N = subject is male
HRT_DUR	Duration of current hormone therapy use (years)	closest to blood draw	continuous numeric .N = subject is male, or current use status is "no" or unknown
HRT_ESTALONE	Any use of estrogen alone	closest to blood draw	0 = never 1 = ever .N = subject is male
HRT_PROGESTIN	Any use of progestin (alone or in combination with estrogen)	closest to blood draw	0 = never 1 = ever .N = subject is male



HRT_YR	Year of data collection (HRT data)	closest to blood draw	YYYY .N = subject is male
<b>ORAL CONTRACEPTIVE USE (AT TIME OF BLOOD DRAW)</b>			
OC	Oral contraceptive use	closest to blood draw	If collected as never/ever: 0 = never 1 = ever If collected as current use (no/yes): 2 = no current use 3 = current use If collected as never/former/current: 4 = never 5 = former 6 = current .N = subject is male
OC_DUR	Duration of current oral contraceptive use (years)	closest to blood draw	continuous numeric if available; if collected as categorical, report the range as free text .N = subject is male, or current use status is "no" or unknown
OC_YR	Year of data collection (OC data)	closest to blood draw	YYYY .N = subject is male
<b>GENOME-WIDE INFORMATION</b>			
GENOMEWIDE_YN	Genomewide data available	most recent	0 = no 1 = yes
GENOME_CHIP	If Genomewide available, what chip? (e.g. "illumina 317k")	most recent	free text

**Appendix 3-** Supplementary table 1 for Article 1

Linear regression for plasma vitamin B2, B6, Folate, B12, Homocysteine and Methionine in relation to lifestyle factors, to nutrients and to food groups as assessed by food frequency questionnaires.

Variable	ratio concentration <sup>a</sup>			ratio concentration <sup>a</sup>			ratio concentration <sup>a</sup>		
	n	(95% CI)	P	n	(95% CI)	P	n	(95% CI)	P
	Serum riboflavin			Serum vitamin B6			Serum folate		
Never smokers	539	1.00 [reference]		539	1.00 [reference]		539	1.00 [reference]	
Former smokers	489	0.96 (0.89 - 1.04)	0.39	489	0.94 (0.87 - 1.01)	0.12	489	0.95 (0.89 - 1.03)	0.25
Current smokers	456	0.83 (0.77 - 0.90)	1x10 <sup>-5</sup>	456	0.86 (0.80 - 0.93)	3x10 <sup>-4</sup>	456	0.87 (0.81 - 0.94)	4x10 <sup>-4</sup>
log2-cotinine <sup>b</sup>	456	0.93 (0.90 - 0.96)	2x10 <sup>-7</sup>	456	0.93 (0.91 - 0.96)	1x10 <sup>-5</sup>	456	0.94 (0.91 - 0.96)	1x10 <sup>-5</sup>
Alcohol intake at recruitment <sup>c</sup>	995	0.96 (0.93 - 1.00)	0.001	995	1.05 (1.02 - 1.08)	7x10 <sup>-4</sup>	995	0.99 (0.96 - 1.03)	0.31
<b>Following variables adjusted for cotinine (in quartiles) and alcohol at recruitment, among all controls</b>									
Vitamin B2 (nutrient)	995	1.22 (1.12 - 1.32)	2x10 <sup>-7</sup>	995	1.01 (0.94 - 1.09)	0.62	995	1.07 (1.00 - 1.16)	0.06
Vitamin B6 (nutrient)	995	1.11 (1.00 - 1.22)	0.04	995	1.10 (1.00 - 1.20)	0.04	995	1.10 (1.00 - 1.20)	0.04
Folate (nutrient)	995	1.1 (1.01 - 1.21)	0.01	995	1.13 (1.04 - 1.23)	0.004	995	1.26 (1.17 - 1.37)	2x10 <sup>-8</sup>
Vitamin B12 (nutrient)	995	1.06 (1.00 - 1.12)	0.01	995	0.98 (0.94 - 1.03)	0.57	995	1.00 (0.95 - 1.05)	0.85
Vegetables	995	1.01 (0.97 - 1.06)	0.87	995	1.06 (1.02 - 1.10)	0.006	995	1.10 (1.06 - 1.15)	3x10 <sup>-5</sup>
Leafy vegetables	959	1.02 (0.99 - 1.04)	0.08	959	1.03 (1.00 - 1.05)	0.02	959	1.05 (1.02 - 1.07)	2x10 <sup>-5</sup>
Fruiting vegetables	992	1.03 (1.00 - 1.06)	0.22	992	1.05 (1.02 - 1.08)	0.003	992	1.05 (1.02 - 1.08)	4x10 <sup>-4</sup>
Root vegetables	950	0.98 (0.96 - 1.01)	0.30	950	1.02 (1.00 - 1.05)	0.18	950	1.02 (1.00 - 1.05)	0.42
Dairy products	992	1.08 (1.05 - 1.12)	5x10 <sup>-7</sup>	992	1.00 (0.97 - 1.03)	0.57	992	0.99 (0.96 - 1.02)	0.54
Meat products	995	1.00 (0.97 - 1.03)	0.75	995	0.98 (0.95 - 1.01)	0.14	995	0.93 (0.91 - 0.96)	1x10 <sup>-5</sup>
	Serum vitamin B12			Serum homocysteine			Serum methionine		
Never smokers	539	1.00 [reference]		539	1.00 [reference]		539	1.00 [reference]	
Former smokers	489	1.06 (1.01 - 1.11)	0.02	489	1.02 (0.98 - 1.07)	0.33	489	0.98 (0.95 - 1.01)	0.21
Current smokers	456	1.00 (0.95 - 1.05)	0.85	456	1.08 (1.04 - 1.13)	0.001	456	0.95 (0.92 - 0.98)	0.002
log2-cotinine <sup>b</sup>	456	0.99 (0.97 - 1.00)	0.12	456	1.03 (1.01 - 1.05)	0.001	456	0.98 (0.96 - 0.99)	9x10 <sup>-5</sup>
Alcohol intake at recruitment <sup>c</sup>	995	1.01 (0.99 - 1.03)	0.72	995	1.03 (1.01 - 1.05)	0.003	995	1.00 (0.99 - 1.02)	0.09
<b>Following variables adjusted for cotinine (in quartiles) and alcohol at recruitment, among all controls</b>									
Vitamin B2 (nutrient)	995	1.09 (1.04 - 1.14)	2x10 <sup>-4</sup>	995	0.88 (0.85 - 0.92)	3x10 <sup>-10</sup>	995	1.04 (1.00 - 1.07)	0.02
Vitamin B6 (nutrient)	995	1.05 (0.99 - 1.11)	0.06	995	0.88 (0.84 - 0.93)	2x10 <sup>-7</sup>	995	1.05 (1.01 - 1.10)	0.003
Folate (nutrient)	995	1.02 (0.97 - 1.08)	0.21	995	0.89 (0.86 - 0.93)	4x10 <sup>-8</sup>	995	1.04 (1.01 - 1.08)	0.01
Vitamin B12 (nutrient)	995	1.08 (1.04 - 1.11)	2x10 <sup>-7</sup>	995	0.92 (0.90 - 0.95)	1x10 <sup>-8</sup>	995	1.02 (1.00 - 1.04)	0.04
Vegetables	995	1.02 (0.99 - 1.05)	0.13	995	0.96 (0.94 - 0.98)	2x10 <sup>-4</sup>	995	1.01 (1.00 - 1.03)	0.09
Leafy vegetables	959	1.01 (1.00 - 1.03)	0.09	959	0.99 (0.98 - 1.00)	0.11	959	1.00 (0.99 - 1.01)	0.72
Fruiting vegetables	992	0.99 (0.97 - 1.01)	0.23	992	0.98 (0.97 - 1.00)	0.02	992	1.00 (0.99 - 1.02)	0.4
Root vegetables	950	1.01 (1.00 - 1.03)	0.06	950	0.99 (0.98 - 1.00)	0.40	950	1.00 (0.99 - 1.01)	0.92
Dairy products	992	1.04 (1.02 - 1.06)	8x10 <sup>-5</sup>	992	0.97 (0.96 - 0.99)	9x10 <sup>-4</sup>	992	1.01 (0.99 - 1.02)	0.16
Meat products	995	1.03 (1.01 - 1.05)	0.002	995	0.96 (0.94 - 0.97)	1x10 <sup>-6</sup>	995	1.01 (1.00 - 1.02)	0.07

<sup>a</sup> Ratios of plasma levels were calculated as 2 raised to the power of the beta estimates from linear regression models with log2- plasma levels as response. The linear regression models were adjusted for case-control status, age, sex, country and as indicated adjusted for cotinine (in quartiles) and alcohol at recruitment. The base 2 logarithm of each nutrient or food group was included as covariate in the linear regression models. The level ratio can thus be interpreted as the average relative change in serum levels that would be expected for a doubling in dietary intake.

<sup>b</sup> Among current smokers

<sup>c</sup> coded 0/1/2/3/4/5/6/ ~ alcohol[g/day]=0 / 0.1-6 / 6.1-12 / 12.1-24 / 24.1-60 / 60.1-96 in men or >60 in women / >90 in men. Adjusted for case-control status, age, sex, country and cotinine (in quartiles) among all controls.

## Appendix 4- Supplementary table 2 for Article 1

Spearman correlation coefficient between plasma vitamin B2, B6, folate, B12, homocysteine and methionine

<b>Overall</b>					
	Pyridoxal 5'-phosphate	Folate	Cobalamin	Homocysteine	Methionine
Riboflavin (Vitamin B2)	0.25**	0.35**	0.09*	-0.12**	0.005
Pyridoxal 5'-phosphate (Vitamin B6)		0.34**	0.14**	-0.14**	0.11**
Folate (vitamin B9)			0.11**	-0.32**	-0.04*
Cobalamin (Vitamin B12)				-0.29**	0.05
Homocysteine					-0.04
<b>Matched controls</b>					
	Pyridoxal 5'-phosphate	Folate	Cobalamin	Homocysteine	Methionine
Riboflavin (Vitamin B2)	0.21**	0.30**	0.08	-0.03	-0.02
Pyridoxal 5'-phosphate (Vitamin B6)		0.30**	0.13*	-0.09**	0.18**
Folate (vitamin B9)			0.15**	-0.22**	-0.01
Cobalamin (Vitamin B12)				-0.25**	0.10*
Homocysteine					-0.09*
<b>Cases</b>					
	Pyridoxal 5'-phosphate	Folate	Cobalamin	Homocysteine	Methionine
Riboflavin (Vitamin B2)	0.29**	0.39**	0.10*	-0.19**	0.03
Pyridoxal 5'-phosphate (Vitamin B6)		0.38**	0.14*	-0.18**	0.04
Folate (vitamin B9)			0.08*	-0.37**	-0.10*
Cobalamin (Vitamin B12)				-0.32**	0.01
Homocysteine					0.00

\* indicates  $0.001 \leq p < 0.05$  and \*\* indicates  $p < 0.001$

## Appendix 5-Supplementary table 3 for Article 1

Hazard ratios and 95% Confidence Intervals of all-cause mortality among Head and Neck Cancer (HNC) and Esophagus Squamous Cell Carcinoma (ESCC) cases for plasma levels of Vitamins B2, B6, Folate, B12, Homocysteine and Methionine

Quartile (Range)	Decesaed <sup>b</sup>	Alive <sup>b</sup>	Head and neck cancer <sup>a</sup>		Esophagus Squamous Cell Carcinoma	
			Minimally adjusted <sup>c</sup>	Adjusted for risk factor <sup>d</sup>	Minimally adjusted <sup>e</sup>	Adjusted for risk factor <sup>f</sup>
<b>Riboflavin (Vitamin B2), nmol/L<sup>g</sup></b>						
1 (2.5 - 9.4)	75 (27.1%)	85 (31.3%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (9.5 - 13.1)	54 (19.5%)	47 (17.3%)	1.26 (0.78 - 2.03)	1.32 (0.81 - 2.16)	0.76 (0.28 - 2.09)	0.59 (0.18 - 1.97)
3 (13.2 - 21.3)	81 (29.2%)	76 (27.9%)	1.07 (0.69 - 1.65)	1.13 (0.72 - 1.78)	0.83 (0.32 - 2.12)	1.07 (0.34 - 3.32)
4 (21.4 - 199)	67 (24.2%)	64 (23.5%)	1.05 (0.62 - 1.77)	1.30 (0.74 - 2.28)	0.85 (0.33 - 2.18)	1.03 (0.33 - 3.25)
P for trend <sup>h</sup>			0.76	0.29	0.92	0.55
<b>Pyridoxal 5'-phosphate (Vitamin B6), nmol/L<sup>g</sup></b>						
1 (7.2 - 25.6)	80 (28.9%)	85 (31.3%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (25.7 - 34.5)	59 (21.3%)	66 (24.3%)	0.80 (0.50 - 1.27)	0.81 (0.49 - 1.32)	2.02 (0.76 - 5.36)	3.65 (1.13 - 11.81)
3 (34.6 - 47.6)	64 (23.1%)	61 (22.4%)	0.89 (0.56 - 1.40)	0.91 (0.56 - 1.48)	1.78 (0.67 - 4.71)	3.26 (1.02 - 10.46)
4 (47.7 - 272)	74 (26.7%)	60 (22.1%)	0.96 (0.60 - 1.53)	0.96 (0.59 - 1.56)	1.73 (0.71 - 4.19)	2.66 (0.96 - 7.39)
P for trend <sup>h</sup>			0.87	0.87	0.51	0.26
<b>Serum folate (vitamin B9), nmol/L<sup>g</sup></b>						
1 (0.3 - 9.1)	73 (26.4%)	104 (38.2%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (9.2 - 12.8)	64 (23.1%)	52 (19.1%)	1.37 (0.85 - 2.21)	1.40 (0.86 - 2.27)	1.16 (0.48 - 2.81)	1.83 (0.68 - 4.89)
3 (12.9 - 18.1)	71 (25.6%)	53 (19.5%)	1.43 (0.90 - 2.26)	1.40 (0.87 - 2.25)	0.77 (0.30 - 2.01)	1.20 (0.41 - 3.51)
4 (18.2 - 109)	69 (24.9%)	63 (23.2%)	0.98 (0.58 - 1.66)	1.01 (0.58 - 1.77)	1.06 (0.42 - 2.69)	1.11 (0.44 - 2.83)
P for trend <sup>h</sup>			0.97	0.98	0.78	0.91
<b>Cobalamin (Vitamin B12), pmol/L<sup>g</sup></b>						
1 (75.1 - 265)	84 (30.3%)	70 (25.7%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (266 - 328)	65 (23.5%)	71 (26.1%)	0.87 (0.56 - 1.34)	0.90 (0.57 - 1.41)	1.24 (0.51 - 3.01)	0.91 (0.33 - 2.52)
3 (329 - 391)	49 (17.7%)	53 (19.5%)	0.74 (0.45 - 1.21)	0.73 (0.44 - 1.21)	0.96 (0.39 - 2.39)	0.66 (0.23 - 1.86)
4 (392 - 2737)	79 (28.5%)	78 (28.7%)	0.60 (0.38 - 0.96)	0.55 (0.34 - 0.90)	1.74 (0.78 - 3.87)	1.18 (0.45 - 3.08)
P for trend <sup>h</sup>			0.06	0.06	0.17	0.49
<b>Homocysteine, µmol/L<sup>g</sup></b>						
1 (4.9 - 8.3)	53 (19.1%)	66 (24.3%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (8.4 - 10.1)	62 (22.4%)	68 (25.0%)	1.00 (0.59 - 1.70)	1.01 (0.59 - 1.74)	0.93 (0.38 - 2.25)	0.62 (0.23 - 1.67)
3 (10.2 - 12.4)	58 (20.9%)	46 (16.9%)	1.39 (0.82 - 2.34)	1.36 (0.80 - 2.32)	1.27 (0.49 - 3.28)	0.98 (0.34 - 2.81)
4 (12.5 - 64.9)	104 (37.5%)	92 (33.8%)	1.53 (0.94 - 2.50)	1.42 (0.85 - 2.37)	1.19 (0.52 - 2.74)	1.03 (0.40 - 2.63)
P for trend <sup>h</sup>			0.44	0.62	0.39	0.25
<b>Methionine, µmol/L<sup>g</sup></b>						
1 (12.6 - 21.8)	83 (30.0%)	72 (26.5%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (21.9 - 25.1)	61 (22.0%)	58 (21.3%)	0.84 (0.54 - 1.32)	0.85 (0.54 - 1.35)	0.80 (0.31 - 2.07)	0.75 (0.28 - 2.06)
3 (25.2 - 29.3)	67 (24.2%)	74 (27.2%)	0.64 (0.39 - 1.03)	0.66 (0.41 - 1.09)	0.93 (0.39 - 2.22)	1.12 (0.40 - 3.14)
4 (29.4 - 62.9)	66 (23.8%)	68 (25.0%)	0.69 (0.44 - 1.11)	0.73 (0.45 - 1.17)	1.89 (0.88 - 4.04)	3.22 (1.31 - 7.94)
P for trend <sup>h</sup>			0.01	0.02	0.56	0.15

<sup>a</sup> Adenocarcinoma excluded

<sup>b</sup> Vital status among cancer cases of the head and neck and esophagus at the last follow-up.

<sup>c</sup> Assessed by analysing HNC cases by Cox-proportional hazards regression, adjusting for country, sex, and age at diagnosis.

<sup>d</sup> Assessed by analysing HNC cases by Cox-proportional hazards regression, adjusting for country, sex, and age at diagnosis, and further by educational attainment (5 groups), smoking status (never/former/current/missing), plasma cotinine (quartiles defined among current smokers) and alcohol intake at recruitment (g/day).

<sup>e</sup> Assessed by analysing ESCC cases by Cox-proportional hazards regression, adjusting for country, sex, and age at diagnosis.

<sup>f</sup> Assessed by analysing ESCC cases by Cox-proportional hazards regression, adjusting for country, sex, and age at diagnosis, and further by educational attainment (5 groups), smoking status (never/former/current/missing), plasma cotinine (quartiles defined among current smokers) and alcohol intake at recruitment (g/day).

<sup>g</sup> Quartile cut-off points were determined based on the plasma level distribution of each biomarker among 516 individually matched controls of the head and neck and esophagus.

<sup>h</sup> P for trend assessed by the base 2 logarithm of the serum levels

Appendix 6- Supplementary table 4 for Article 1

Odds ratios of Head and Neck Cancer (HNC) and Esophagus Squamous Cell Carcinoma (ESCC) for plasma levels of Vitamins B2, B6, Folate, B12, homocysteine and methionine excluding cases that were diagnosed with HNC and ESCC cancer less than a year after blood draw.

Quartile (Range)	Case <sup>b</sup>	Control <sup>b</sup>	Odds Ratio (95% Confidence Interval)					
			Head and neck cancer <sup>a</sup>			Esophagus Squamous Cell Carcinoma		
			Matched controls unadjusted <sup>c</sup> (n=323/323) <sup>e</sup>	Matched controls adjusted <sup>d</sup> (n=323/323) <sup>d</sup>	All controls combined adjusted <sup>e</sup> (n = 323/923) <sup>e</sup>	Matched controls unadjusted <sup>f</sup> (n=68/68) <sup>f</sup>	Matched controls adjusted <sup>d</sup> (n=68/68) <sup>d</sup>	All controls combined adjusted <sup>e</sup> n = (68/923) <sup>e</sup>
<b>Riboflavin (Vitamin B2), nmol/L<sup>h</sup></b>								
1 (2.5 - 9.4)	140 (29.2%)	122 (25.5%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (9.5 - 13.1)	90 (18.8%)	115 (24.0%)	0.66 (0.42 - 1.03)	0.83 (0.48 - 1.43)	0.81 (0.54 - 1.21)	0.51 (0.17 - 1.54)	0.31 (0.03 - 2.76)	0.51 (0.21 - 1.25)
3 (13.2 - 21.3)	135 (28.2%)	119 (24.8%)	1.07 (0.69 - 1.66)	1.50 (0.88 - 2.55)	1.17 (0.79 - 1.73)	0.68 (0.25 - 1.88)	0.42 (0.05 - 3.41)	0.96 (0.44 - 2.06)
4 (21.4 - 199)	114 (23.8%)	123 (25.7%)	0.67 (0.42 - 1.07)	1.20 (0.67 - 2.15)	1.22 (0.78 - 1.89)	0.73 (0.26 - 2.04)	1.14 (0.19 - 6.73)	1.17 (0.51 - 2.69)
<i>P</i> for trend <sup>i</sup>			0.21	0.22	0.29	0.43	0.69	0.46
<b>Pyridoxal 5'-phosphate (Vitamin B6), nmol/L<sup>h</sup></b>								
1 (7.2 - 25.6)	150 (31.3%)	122 (25.5%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (25.7 - 34.5)	106 (22.1%)	115 (24.0%)	0.72 (0.46 - 1.12)	0.85 (0.5 - 1.44)	0.97 (0.65 - 1.44)	0.86 (0.32 - 2.31)	1.51 (0.34 - 6.64)	0.98 (0.42 - 2.28)
3 (34.6 - 47.6)	104 (21.7%)	119 (24.8%)	0.68 (0.44 - 1.06)	0.99 (0.57 - 1.7)	1.10 (0.73 - 1.66)	0.92 (0.32 - 2.60)	0.93 (0.17 - 5.11)	1.28 (0.55 - 2.98)
4 (47.7 - 272)	119 (24.8%)	123 (25.7%)	0.75 (0.48 - 1.16)	0.98 (0.57 - 1.7)	1.01 (0.67 - 1.53)	1.17 (0.46 - 2.94)	1.96 (0.40 - 9.48)	2.36 (1.08 - 5.18)
<i>P</i> for trend <sup>i</sup>			0.23	0.83	0.68	0.86	0.72	0.06
<b>Serum folate (vitamin B9), nmol/L<sup>h</sup></b>								
1 (0.3 - 9.1)	152 (31.7%)	117 (24.4%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (9.2 - 12.8)	101 (21.1%)	121 (25.3%)	0.50 (0.32 - 0.81)	0.63 (0.35 - 1.13)	0.81 (0.54 - 1.21)	0.89 (0.31 - 2.54)	5.09 (0.72 - 35.75)	1.05 (0.49 - 2.28)
3 (12.9 - 18.1)	106 (22.1%)	118 (24.6%)	0.58 (0.36 - 0.96)	0.73 (0.40 - 1.33)	1.09 (0.73 - 1.65)	0.61 (0.20 - 1.84)	1.92 (0.30 - 12.34)	0.94 (0.40 - 2.20)
4 (18.2 - 109)	120 (25.1%)	123 (25.7%)	0.47 (0.28 - 0.79)	0.62 (0.33 - 1.16)	0.93 (0.61 - 1.42)	0.96 (0.30 - 3.09)	6.71 (0.63 - 71.05)	0.90 (0.40 - 2.01)
<i>P</i> for trend <sup>i</sup>			2x10 <sup>-4</sup>	0.009	0.32	0.66	0.27	0.97
<b>Cobalamin (Vitamin B12), pmol/L<sup>h</sup></b>								
1 (75.1 - 265)	139 (29.0%)	119 (24.8%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (266 - 328)	121 (25.3%)	120 (25.1%)	0.79 (0.51 - 1.23)	1.01 (0.59 - 1.72)	0.83 (0.56 - 1.23)	1.20 (0.39 - 3.74)	1.17 (0.23 - 5.90)	1.08 (0.49 - 2.36)
3 (329 - 391)	86 (18.0%)	120 (25.1%)	0.56 (0.34 - 0.91)	0.59 (0.32 - 1.07)	0.64 (0.42 - 0.98)	0.80 (0.26 - 2.42)	1.41 (0.31 - 6.32)	0.92 (0.41 - 2.07)
4 (392 - 2737)	133 (27.8%)	120 (25.1%)	0.86 (0.53 - 1.39)	1.06 (0.59 - 1.90)	0.89 (0.60 - 1.31)	1.16 (0.36 - 3.75)	1.02 (0.21 - 4.94)	1.05 (0.49 - 2.27)
<i>P</i> for trend <sup>i</sup>			0.89	0.78	0.70	0.88	0.73	0.86
<b>Homocysteine, µmol/L<sup>h</sup></b>								
1 (4.9 - 8.3)	101 (21.1%)	122 (25.5%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (8.4 - 10.1)	114 (23.8%)	122 (25.5%)	1.40 (0.86 - 2.25)	1.47 (0.83 - 2.60)	1.33 (0.88 - 2.03)	0.9 (0.31 - 2.62)	0.62 (0.12 - 3.13)	1.10 (0.48 - 2.54)
3 (10.2 - 12.4)	88 (18.4%)	117 (24.4%)	1.15 (0.70 - 1.89)	1.06 (0.59 - 1.94)	0.91 (0.58 - 1.44)	1.23 (0.42 - 3.58)	1.05 (0.23 - 4.79)	0.84 (0.34 - 2.05)
4 (12.5 - 64.9)	176 (36.7%)	118 (24.6%)	3.21 (1.84 - 5.62)	2.61 (1.32 - 5.17)	1.89 (1.22 - 2.91)	1.72 (0.64 - 4.64)	0.99 (0.23 - 4.24)	1.85 (0.80 - 4.29)
<i>P</i> for trend <sup>i</sup>			7x10 <sup>-5</sup>	0.007	0.003	0.06	0.53	0.07
<b>Methionine, µmol/L<sup>h</sup></b>								
1 (12.6 - 21.8)	138 (28.8%)	121 (25.3%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (21.9 - 25.1)	103 (21.5%)	118 (24.6%)	0.74 (0.50 - 1.12)	0.96 (0.59 - 1.56)	0.86 (0.58 - 1.29)	0.87 (0.35 - 2.19)	4.18 (0.48 - 36.12)	0.69 (0.32 - 1.50)
3 (25.2 - 29.3)	118 (24.6%)	120 (25.1%)	0.86 (0.55 - 1.34)	1.21 (0.7 - 2.11)	0.92 (0.61 - 1.38)	0.57 (0.20 - 1.59)	0.78 (0.13 - 4.84)	0.87 (0.40 - 1.92)
4 (29.4 - 62.9)	120 (25.1%)	120 (25.1%)	0.91 (0.56 - 1.46)	1.4 (0.77 - 2.54)	1.06 (0.70 - 1.61)	0.60 (0.19 - 1.85)	1.50 (0.18 - 12.27)	1.08 (0.49 - 2.38)
<i>P</i> for trend <sup>i</sup>			0.42	0.36	0.94	0.18	0.59	0.90

<sup>a</sup> Adenocarcinoma excluded

<sup>b</sup> Numbers include all cancer cases of the head and neck and esophagus, and individually matched control for whom laboratory measurements were available for complete case sets. Uninformative case-sets were excluded.

<sup>c</sup> Assessed by analysing HNC cases and their individually matched controls by conditional logistic regression, conditioning on individual case set.

<sup>d</sup> Further adjusted for educational attainment (in 5 groups), smoking status (never/former/current/missing), cotinine (quartiles defined among current smokers) and alcohol intake at recruitment(g/day)

<sup>e</sup> Assessed by analysing HNC cases and all controls combined by unconditional logistic regression, adjusting for country, sex, age at recruitment (in 5 year groups), educational attainment (in 5 groups), smoking status (never/former/current/missing), cotinine (quartiles defined among current smokers) and alcohol intake at recruitment (g/day).

<sup>f</sup> Assessed by analysing ESCC cases and their individually matched controls by conditional logistic regression, conditioning on individual case set.

<sup>g</sup> Assessed by analysing ESCC cases and all controls combined by unconditional logistic regression, adjusting for country, sex, age at recruitment (in 5 year groups), educational attainment (in 5 groups), smoking status (never/former/current/missing), cotinine (quartiles defined among current smokers) and alcohol intake at recruitment (g/day).

<sup>h</sup> Quartile cut-off points were determined based on the plasma level distribution of each biomarker among the 516 individually matched controls of the head and neck and esophagus .

<sup>i</sup> *P* for trend assessed by the base 2 logarithm of the serum levels

## Appendix 7- Supplementary table 5 for Article 1

Odds ratios of Head and Neck Cancer (HNC) and Esophagus Squamous Cell Carcinoma (ESCC) for dietary intake of B-vitamins and plasma levels of vitamin B2, B6, Folate and B12.

Quartile (Range)	Case <sup>b</sup>	Control <sup>b</sup>	Odds Ratio (95% Confidence Interval)			
			Head and neck cancer <sup>a</sup>		Esophagus Squamous Cell Carcinoma <sup>a</sup>	
			Overall - by dietary intake <sup>c</sup>	Overall - by serum level <sup>d</sup> from table 2	Overall - by dietary intake <sup>d</sup>	Overall - by serum level <sup>d</sup> from table 2
<b>Riboflavin (Vitamin B2), mg/day<sup>d</sup></b>						
1 (0.52 - 1.32)	145 (28.2%)	129 (24.7%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (1.33 - 1.70)	125 (24.3%)	130 (24.9%)	0.95 (0.59 - 1.53)	0.75 (0.45 - 1.27)	0.14 (0.03 - 0.71)	0.17 (0.02 - 1.26)
3 (1.71 - 2.23)	116 (22.6%)	131 (25.1%)	0.82 (0.50 - 1.35)	1.33 (0.80 - 2.20)	1.06 (0.25 - 4.42)	0.57 (0.09 - 3.56)
4 (2.24 - 4.35)	128 (24.9%)	132 (25.3%)	0.84 (0.48 - 1.46)	1.19 (0.68 - 2.08)	2.29 (0.38 - 13.92)	0.94 (0.18 - 4.92)
P for trend <sup>e</sup>			0.28	0.22	0.60	0.84
<b>Pyridoxal 5'-phosphate (Vitamin B6), mg/day<sup>d</sup></b>						
1 (0.57 - 1.51)	160 (31.1%)	129 (24.7%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (1.52 - 1.89)	96 (18.7%)	130 (24.9%)	0.53 (0.31 - 0.90)	0.75 (0.45 - 1.26)	0.41 (0.11 - 1.56)	1.96 (0.51 - 7.56)
3 (1.90 - 2.36)	127 (24.7%)	131 (25.1%)	0.67 (0.38 - 1.15)	0.96 (0.57 - 1.63)	1.20 (0.20 - 7.22)	1.41 (0.28 - 7.18)
4 (2.37 - 4.43)	131 (25.5%)	132 (25.3%)	0.53 (0.29 - 0.97)	1.01 (0.59 - 1.71)	3.25 (0.48 - 22.11)	1.82 (0.45 - 7.42)
P for trend <sup>e</sup>			0.08	0.48	0.29	0.82
<b>Serum folate (vitamin B9), µg/day<sup>d</sup></b>						
1 (87 - 237)	174 (33.9%)	129 (24.7%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (238 - 291)	100 (19.5%)	130 (24.9%)	0.68 (0.42 - 1.12)	0.62 (0.36 - 1.08)	0.64 (0.13 - 3.21)	7.78 (1.24 - 49.05)
3 (292 - 367)	113 (22%)	131 (25.1%)	0.62 (0.37 - 1.04)	0.77 (0.44 - 1.36)	0.63 (0.15 - 2.61)	4.08 (0.74 - 22.60)
4 (368 - 1057)	127 (24.7%)	132 (25.3%)	0.81 (0.46 - 1.41)	0.62 (0.34 - 1.13)	1.63 (0.24 - 11.01)	12.45 (1.14 - 135.5)
P for trend <sup>e</sup>			0.13	0.02	0.62	0.14
<b>Cobalamin (Vitamin B12), µg/day<sup>d</sup></b>						
1 (0.21 - 4.04)	125 (24.3%)	129 (24.7%)	1[Reference]	1[Reference]	1[Reference]	1[Reference]
2 (4.05 - 5.75)	123 (23.9%)	130 (24.9%)	1.05 (0.61 - 1.81)	1.01 (0.61 - 1.68)	1.52 (0.42 - 5.52)	1.49 (0.33 - 6.73)
3 (5.76 - 7.76)	120 (23.3%)	131 (25.1%)	0.90 (0.52 - 1.56)	0.59 (0.33 - 1.05)	1.17 (0.27 - 5.03)	1.07 (0.27 - 4.32)
4 (7.77 - 28.85)	146 (28.4%)	132 (25.3%)	0.90 (0.51 - 1.58)	1.20 (0.68 - 2.10)	1.00 (0.28 - 3.61)	0.94 (0.21 - 4.25)
P for trend <sup>e</sup>			0.81	0.64	0.82	0.78

<sup>a</sup> Adenocarcinoma excluded

<sup>b</sup> Numbers include all cancer cases of the head and neck and esophagus, and individually matched control for whom food frequency questionnaire measurements were available for complete case sets. Uninformative case-sets were excluded.

<sup>c</sup> Assessed by analysing HNC cancer cases and their individually matched controls by conditional logistic regression, conditioning on individual case set and adjusting by educational attainment (in 5 groups), smoking status (never/former/current/missing), cotinine (quartiles defined among current smokers) and alcohol intake at recruitment.

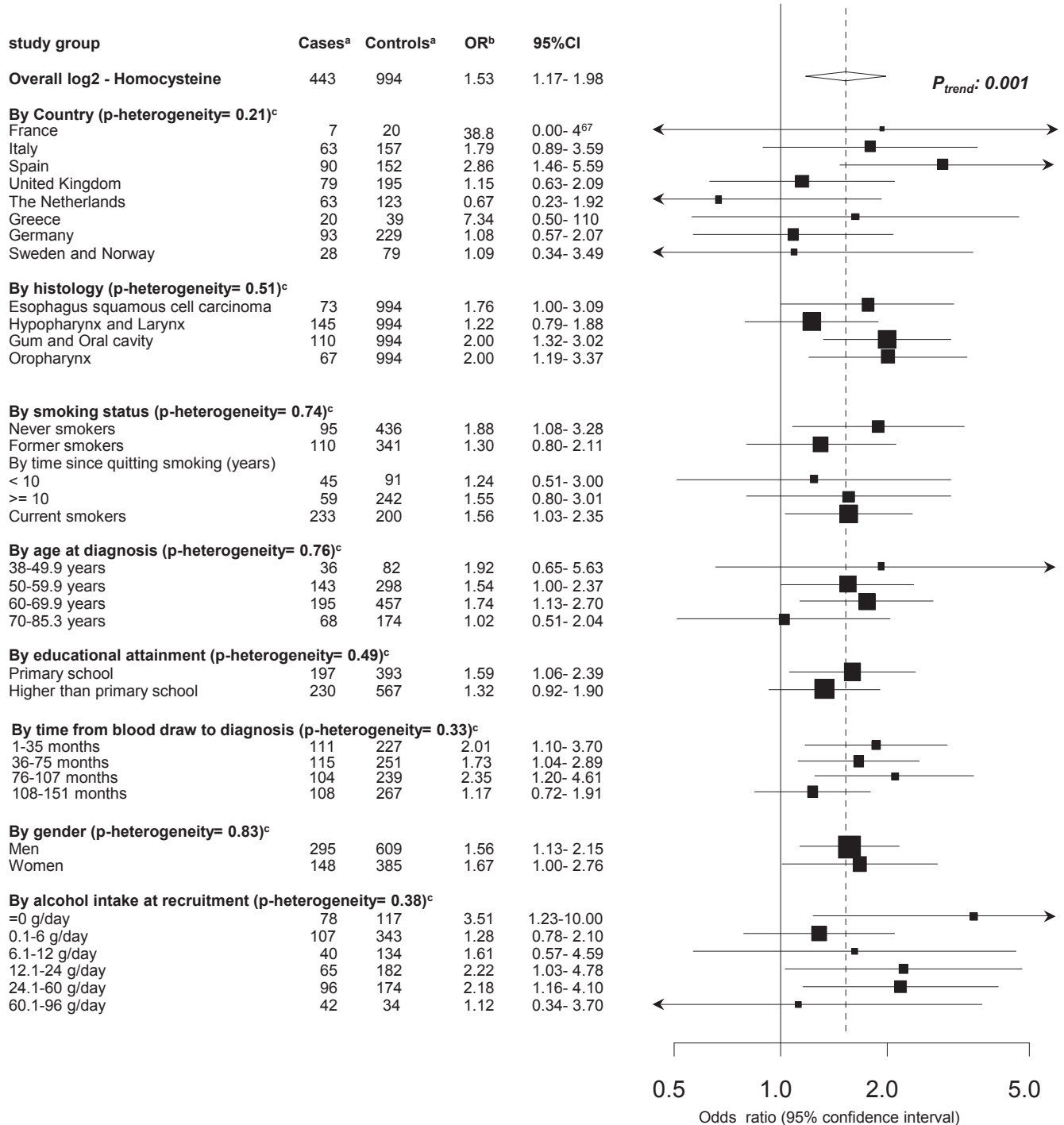
<sup>d</sup> Assessed by analysing ESCC cancer cases and their individually matched controls by conditional logistic regression, conditioning on individual case set and adjusting by educational attainment (in 5 groups), smoking status (never/former/current/missing), cotinine (quartiles defined among current smokers) and alcohol intake at recruitment.

<sup>e</sup> Quartile cut-off points were determined based on the dietary intake distribution of each nutrient among the 516 individually matched controls of the head and neck and esophagus.

<sup>f</sup> P for trend assessed by the base 2 logarithm of the serum levels

Appendix 8- Supplementary figure 1 for Article 1

Forest plot showing Stratified OR of cancer of the head and neck and esophagus for log<sub>2</sub>- of plasma homocysteine.



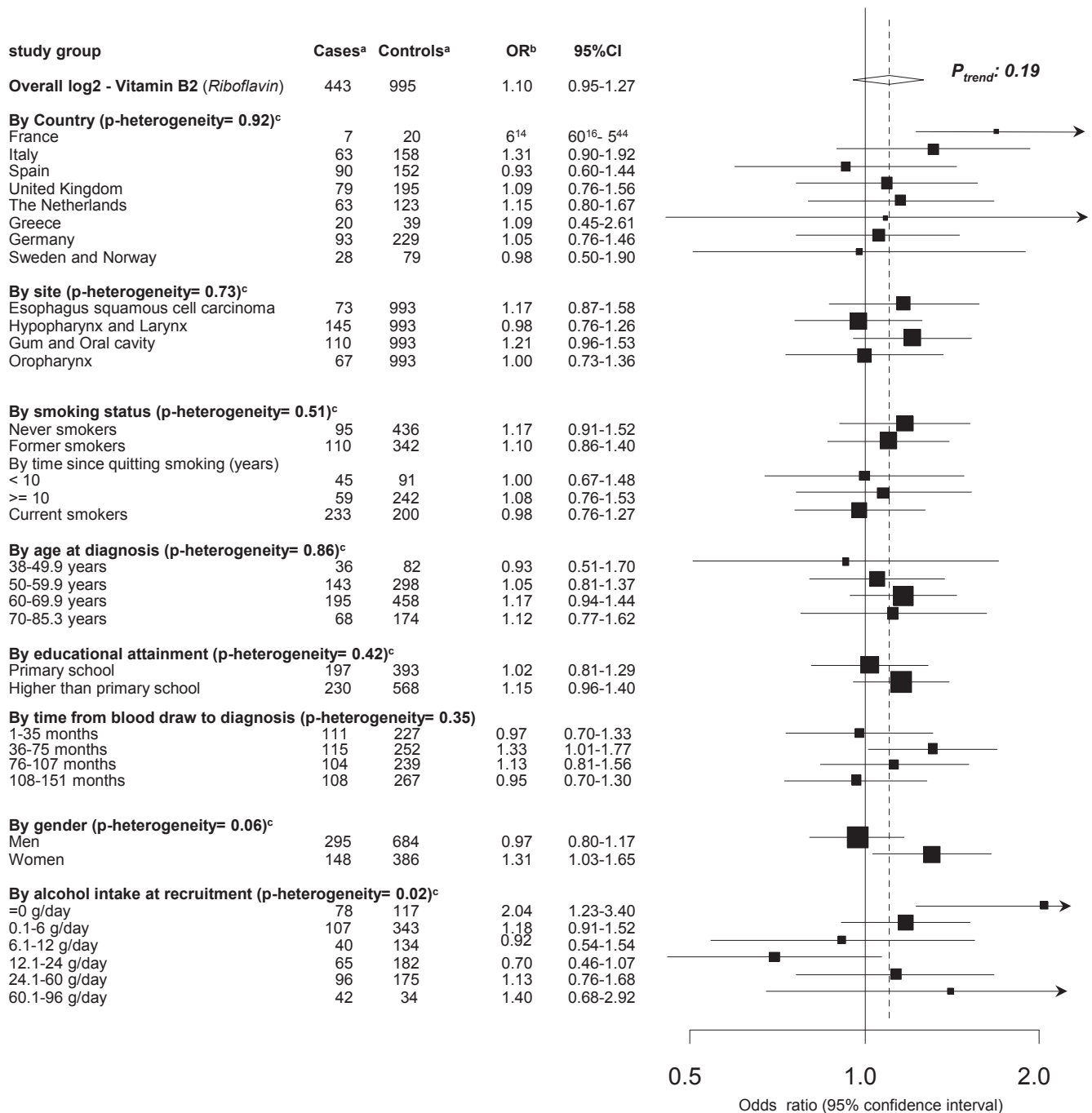
<sup>a</sup> Head and neck and esophagus cancer cases (adenocarcinoma excluded) and controls included in each stratified analysis (control group 2 were included).

<sup>b</sup> ORs were assessed by unconditional logistic regression by including the base 2 logarithm of plasma biomarker levels (ORs indicate relative risks of a doubling in plasma levels), and where relevant adjusted for age, sex, and country; the black dots indicate the ORs and the horizontal lines indicate the 95% confidence intervals.

<sup>c</sup> P heterogeneity indicates results of chi-square test assessing the null hypothesis of ORs being the identical.

Appendix 9- Supplementary figure 2 for Article 1

Forest plot showing Stratified OR of cancer of the head and neck and esophagus for log<sub>2</sub>- of plasma homocysteine vitamin B2.



<sup>a</sup> Head and neck and esophagus cancer cases (adenocarcinoma excluded) and controls included in each stratified analysis (control group 2 were included).

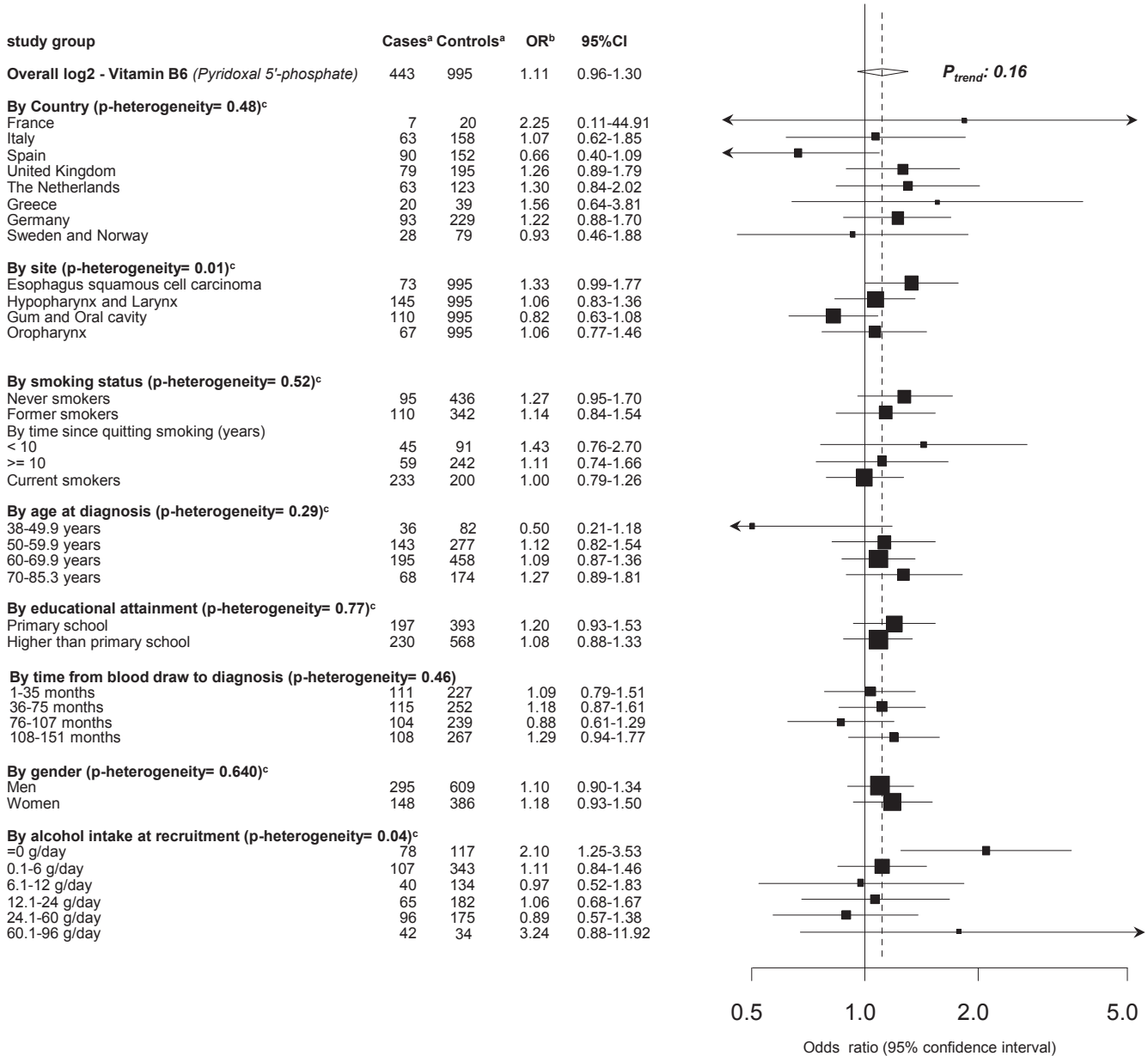
<sup>b</sup> ORs were assessed by unconditional logistic regression by including the base 2 logarithm of plasma biomarker levels (ORs indicate relative risks of a doubling in plasma levels), and where relevant adjusted for age, sex, and country; the black dots indicate the ORs and the horizontal lines indicate the 95% confidence intervals.

<sup>c</sup> P heterogeneity indicates results of chi-square test assessing the null hypothesis of ORs being the identical.



Appendix 10- Supplementary figure 3 for Article 1

Forest plot showing Stratified OR of cancer of the head and neck and esophagus for log<sub>2</sub>- of plasma homocysteine vitamin B6.



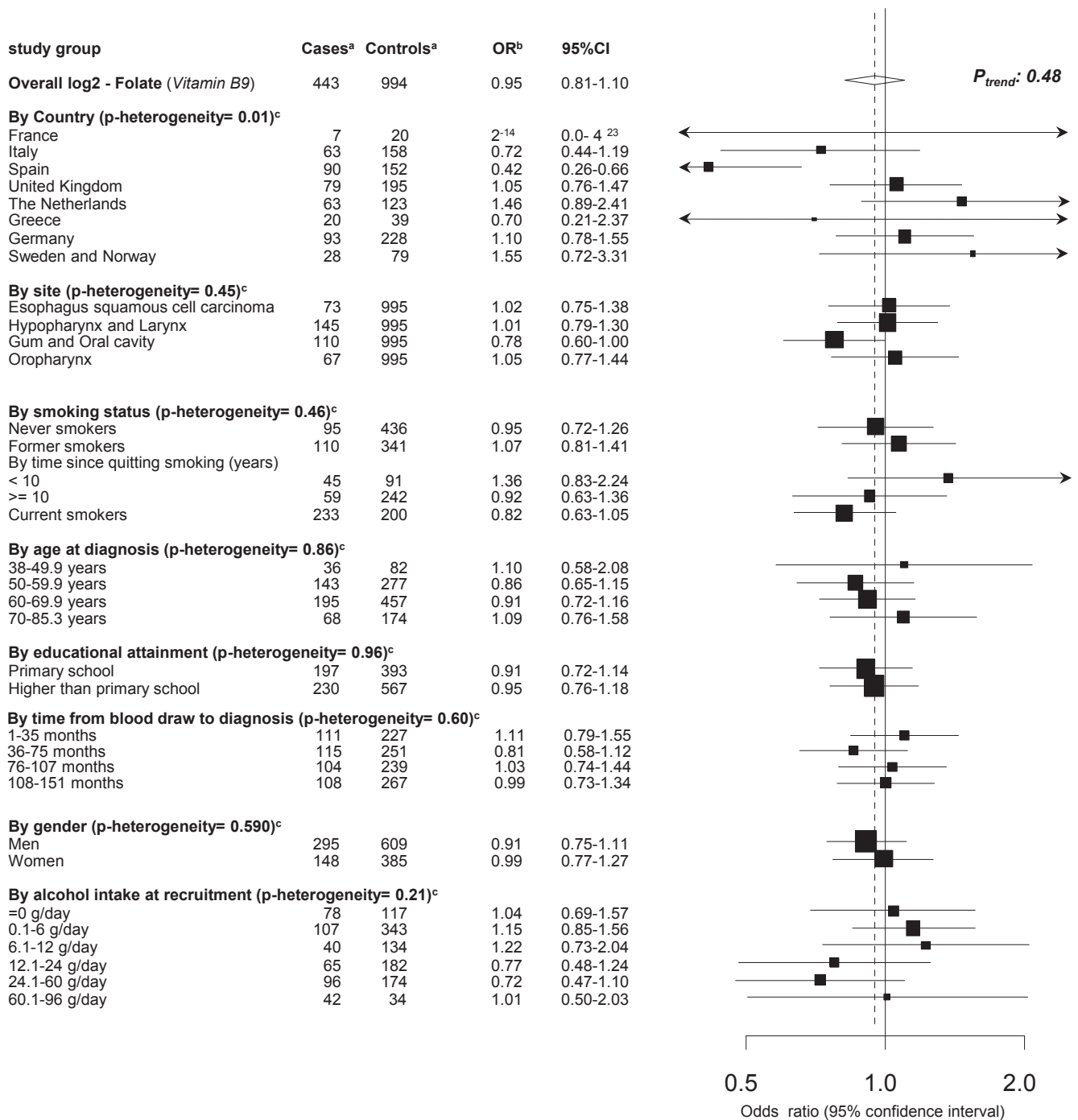
<sup>a</sup> Head and neck and esophagus cancer cases (adenocarcinoma excluded) and controls included in each stratified analysis (control group 2 were included).

<sup>b</sup> ORs were assessed by unconditional logistic regression by including the base 2 logarithm of plasma biomarker levels (ORs indicate relative risks of a doubling in plasma levels), and where relevant adjusted for age, sex, and country; the black dots indicate the ORs and the horizontal lines indicate the 95% confidence intervals.

<sup>c</sup> P heterogeneity indicates results of chi-square test assessing the null hypothesis of ORs being the identical.

Appendix 11- Supplementary figure 4 for Article 1

Forest plot showing Stratified OR of cancer of the head and neck and esophagus for log<sub>2</sub>- of plasma folate.



<sup>a</sup> Head and neck and esophagus cancer cases (adenocarcinoma excluded) and controls included in each stratified analysis (control group 2 were included).

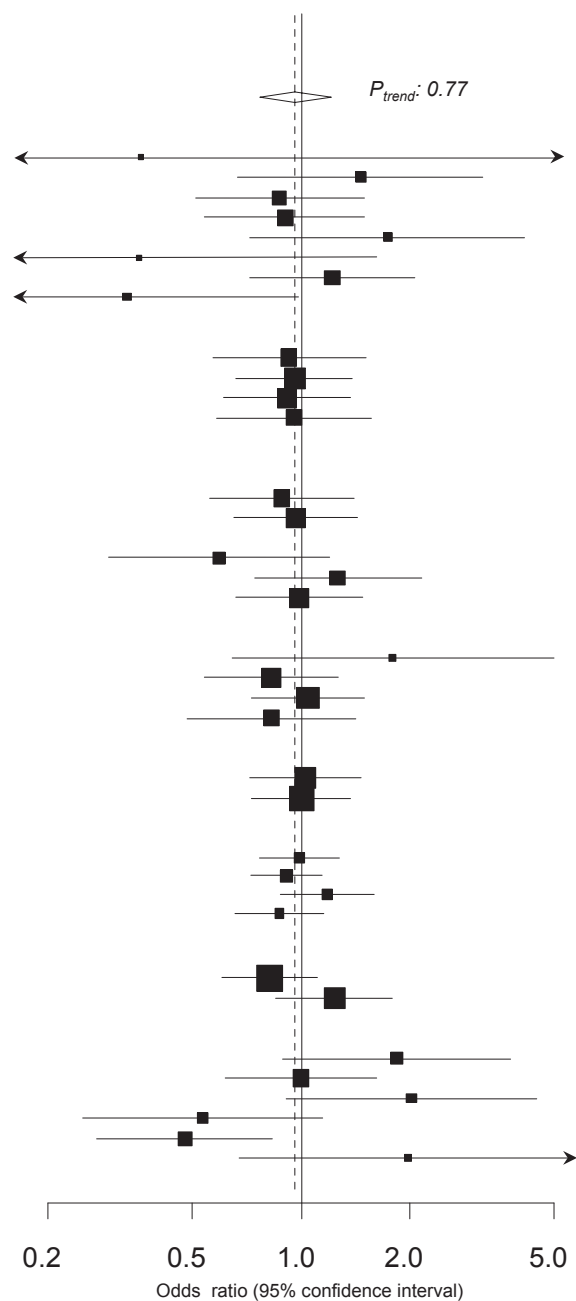
<sup>b</sup> ORs were assessed by unconditional logistic regression by including the base 2 logarithm of plasma biomarker levels (ORs indicate relative risks of a doubling in plasma levels), and where relevant adjusted for age, sex, and country; the black dots indicate the ORs and the horizontal lines indicate the 95% confidence intervals.

<sup>c</sup> P heterogeneity indicates results of chi-square test assessing the null hypothesis of ORs being the identical.

Appendix 12- Supplementary figure 5 for Article 1

Forest plot showing stratified OR of cancer of the head and neck and esophagus for log<sub>2</sub>- of plasma vitamin B12.

study group	Cases <sup>a</sup>	Controls <sup>a</sup>	OR <sup>b</sup>	95%CI
<b>Overall log<sub>2</sub> - Vitamin B12 (Cobalamin )</b>	443	995	0.97	0.768-1.22
<b>By Country (p-heterogeneity= 0.34)<sup>c</sup></b>				
France	7	20	0.26	0.010-6.82
Italy	63	158	1.46	0.668-3.20
Spain	90	152	0.87	0.511-1.49
United Kingdom	79	195	0.90	0.540-1.50
The Netherlands	63	123	1.74	0.725-4.18
Greece	20	39	0.36	0.078-1.61
Germany	93	228	1.22	0.719-2.07
Sweden and Norway	28	79	0.33	0.071-1.52
<b>By site (p-heterogeneity= 0.99)<sup>c</sup></b>				
Esophagus squamous cell carcinoma	73	995	0.93	0.571-1.51
Hypopharynx and Larynx	145	995	0.96	0.660-1.39
Gum and Oral cavity	110	995	0.91	0.607-1.37
Oropharynx	67	995	0.96	0.582-1.57
<b>By smoking status (p-heterogeneity= 0.74)<sup>c</sup></b>				
Never smokers	95	436	0.88	0.559-1.40
Former smokers	110	342	0.97	0.654-1.43
<b>By time since quitting smoking (years)</b>				
< 10	45	91	0.59	0.294-1.20
>= 10	59	242	1.26	0.743-2.15
Current smokers	233	200	0.99	0.661-1.48
<b>By age at diagnosis (p-heterogeneity= 0.66)<sup>c</sup></b>				
38-49.9 years	36	82	1.79	0.644-5.00
50-59.9 years	143	298	0.83	0.541-1.27
60-69.9 years	195	458	1.04	0.725-1.50
70-85.3 years	68	174	0.83	0.481-1.42
<b>By educational attainment (p-heterogeneity= 0.65)<sup>c</sup></b>				
Primary school	197	393	1.03	0.719-1.47
Higher than primary school	230	568	1.00	0.731-1.38
<b>By time from blood draw to diagnosis (p-heterogeneity= 0.46)</b>				
1-35 months	111	227	0.99	0.614-1.61
36-75 months	115	251	0.85	0.557-1.29
76-107 months	104	239	1.39	0.801-2.40
108-151 months	108	267	0.79	0.467-1.33
<b>By gender (p-heterogeneity= 0.09)<sup>c</sup></b>				
Men	295	609	0.82	0.602-1.10
Women	148	386	1.24	0.851-1.80
<b>By alcohol intake at recruitment (p-heterogeneity= 0.02)<sup>c</sup></b>				
=0 g/day	78	117	1.84	0.892-3.80
0.1-6 g/day	107	343	1.00	0.618-1.62
6.1-12 g/day	40	134	2.03	0.912-4.50
12.1-24 g/day	65	182	0.53	0.248-1.15
24.1-60 g/day	96	175	0.48	0.272-0.83
60.1-96 g/day	42	34	2.18	0.481-9.84



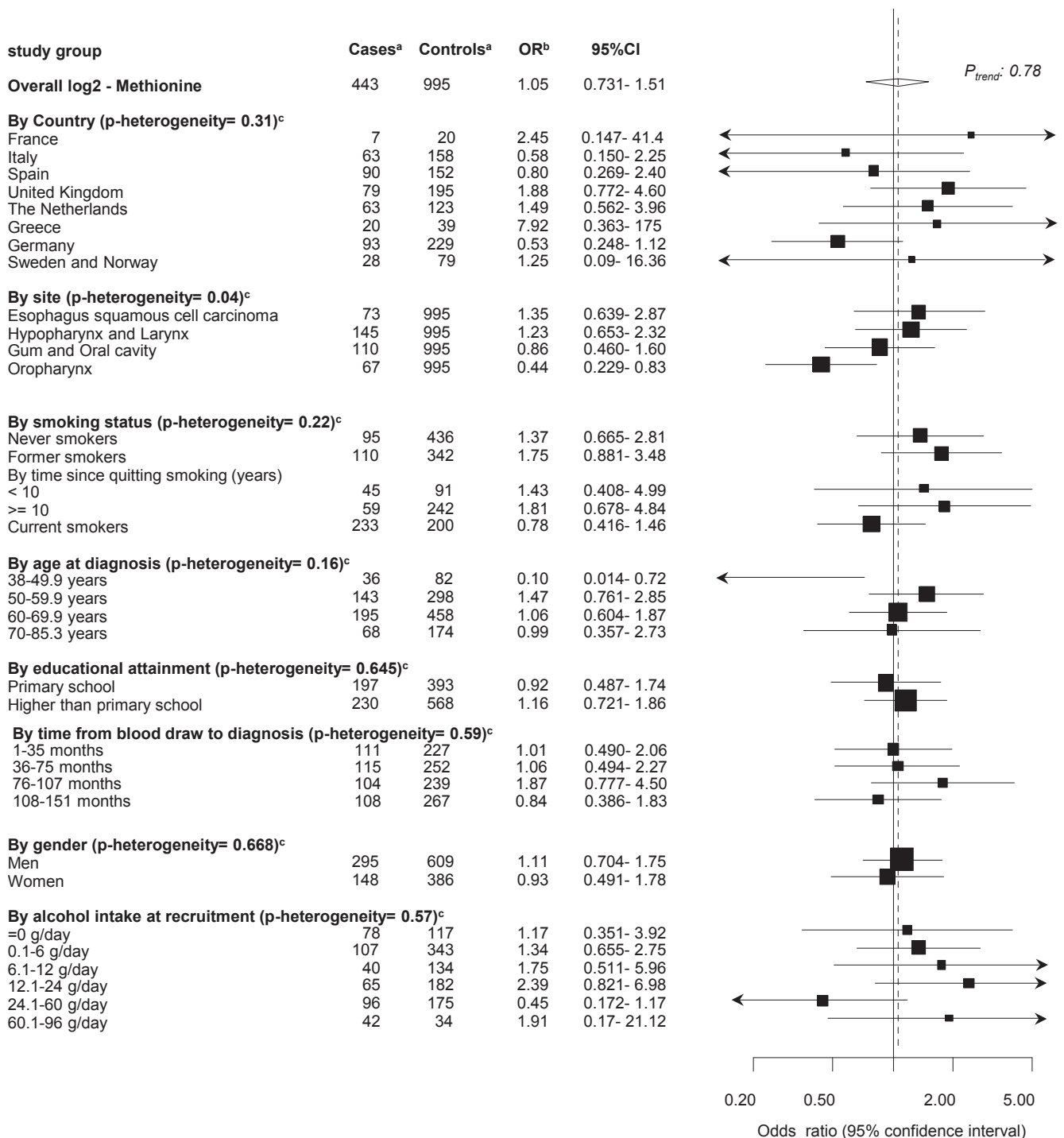
<sup>a</sup> Head and neck and esophagus cancer cases (adenocarcinoma excluded) and controls included in each stratified analysis (control group 2 were included).

<sup>b</sup> ORs were assessed by unconditional logistic regression by including the base 2 logarithm of plasma biomarker levels (ORs indicate relative risks of a doubling in plasma levels), and where relevant adjusted for age, sex, and country; the black dots indicate the ORs and the horizontal lines indicate the 95% confidence intervals.

<sup>c</sup> P heterogeneity indicates results of chi-square test assessing the null hypothesis of ORs being the identical.

**Appendix 13-** Supplementary figure 6 for Article 1

Forest plot showing stratified OR of cancer of the head and neck and esophagus for log<sub>2</sub>- of plasma methionine.



<sup>a</sup> Head and neck and esophagus cancer cases (adenocarcinoma excluded) and controls included in each stratified analysis (control group 2 were included).

<sup>b</sup> ORs were assessed by unconditional logistic regression by including the base 2 logarithm of plasma biomarker levels (ORs indicate relative risks of a doubling in plasma levels), and where relevant adjusted for age, sex, and country; the black dots indicate the ORs and the horizontal lines indicate the 95% confidence intervals.

<sup>c</sup> P heterogeneity indicates results of chi-square test assessing the null hypothesis of ORs being the identical.

## **Appendix 14-** Supplementary methods for Article 1

### *Study cohort – The European Prospective Investigation into Cancer and Nutrition (EPIC)*

521,330 individuals were recruited to the cohort between 1992 and 2000 from 10 European countries, of whom 385,747 donated a blood sample. Blood fractions were aliquoted into 0.5 mL straws, which were heat-sealed and stored in liquid nitrogen tanks at -196°C, except in Umeå, Sweden, where samples were stored in 1.8 mL plastic tubes in -80°C freezers. Participants completed self-administered questionnaires on lifestyle factors and diet. All participants gave written informed consent. The study was approved by the Ethics Committee at the International Agency for Research on Cancer (IARC), Lyon, France, as well as by the local ethics committees of the study centers.

### *Follow-up for cancer incidence and mortality data*

The ascertainment of cases within EPIC reflects the predominant pattern of diagnosis in the general population. Incident cancer cases were identified at regular intervals through population-based cancer registries (Denmark, Italy except Naples, the Netherlands, Norway, Spain, Sweden and United Kingdom) or by active follow-up (France, Germany, Greece and Naples), which involved a combination of methods, including review of health insurance records, cancer and pathology registries, as well as direct contact with participants and their next-of-kin.

Mortality data, including vital status, cause of death, and date of death, were obtained from mortality registries at the regional or national level. Subjects were followed up from study entry until cancer diagnosis (except non-melanoma skin cancer), death, emigration, or the end of the follow-up period for the relevant study center. End of follow-up was defined as the latest date of complete follow-up for both cancer incidence and vital status and varied between study centers from December 2004 to June 2010. Vital status at follow-up is over 98% complete.

### *Validation study – The Melbourne Collaborative Cohort Study (MCCS)*

The MCCS is a prospective cohort study of 41,514 women and men aged 40 to 79 years at the time of recruitment which took place between 1990 and 1994 in Melbourne, Australia. Following blood draw, plasma fractions were stored in liquid nitrogen tanks at -196°C. Incident cancer diagnoses were identified via routine linkage to the Victorian Cancer Registry (VCR). There is a statutory requirement that all cancers diagnosed in the state of Victoria (the state in which Melbourne is situated), excluding non-melanoma skin cancer be reported to the VCR. Vital status and cause of death information were obtained via linkage to the National Death Index of Australia.

### *Biochemical analyses*

Plasma samples were sent on dry ice to the Bevital A/S laboratory (<http://www.bevital.no>) in Bergen, Norway. The study included measurements of plasma concentrations of vitamin B2 (riboflavin), vitamin B6 (measured as pyridoxal 5'-phosphate, its active form), folate (vitamin B9), vitamin B12 (cobalamin), total homocysteine, and methionine. We also measured cotinine as an indicator of recent smoking behavior. Concentrations of B2, B6, homocysteine, methionine, and cotinine were determined by mass spectrometry based methods (liquid chromatography coupled to tandem mass spectrometry; gas chromatography coupled to tandem mass spectrometry),<sup>307-309</sup> and microbiological methods were used to determine concentrations of folate (*Lactobacillus casei*)<sup>310</sup> and B12 (*Lactobacillus leichmannii*).\_ENREF\_285

Samples were analyzed in batches of 86 and quality control included 6 calibration samples, 2 control samples, and 1 blank sample in each batch. The coefficients of variations (CVs) within and between batches were, respectively, 6 and 11% for vitamin B2, 3% and 6% for vitamin B6, 4% and 5% for folate and vitamin B12, 1% and 3% for methionine, and 1% and 2% for homocysteine. All plasma samples were kept at -80°C and all RCC cases and their individual matched controls were analyzed together within the same batches in random order, as were samples from control group 2. The laboratory staff was blinded to the case-control status of the blood samples.

**Appendix 15-** Supplementary table 1 for Article 2

Partial correlation coefficients between circulating concentrations of vitamins B2, B6, folate, B12, homocysteine and methionine, and dietary intake, alcohol intake, circulating cotinine, BMI, and waist-to-hip ratio

	Circulating metabolites					
	Vitamin B6	Vitamin B2	Folate	Vitamin B12	Homocysteine	Methionine
Total vegetable intake	0.07	-0.03	0.05	-0.03	-0.01	0.10
Total fruit, nut, and seed intake	0.17	0.09	0.15	-0.01	-0.08	0.03
Total dairy product intake	0.07	0.19	0.01	0.03	-0.10	0.10
Total meat and meat product intake	-0.06	-0.06	-0.02	-0.02	-0.07	-0.09
Circulating cotinine	-0.25	-0.15	-0.16	-0.03	0.16	-0.12
Alcohol intake at recruitment	0.13	-0.04	0.03	-0.04	0.05	0.00
Body Mass Index	-0.13	0.00	-0.13	-0.06	0.05	-0.05
Waist-to-hip ratio	-0.13	0.05	-0.06	-0.07	0.00	-0.03

a) Calculated as the correlation between the given variables among the 556 matched controls (control group 1) after conditioning on country, age at baseline, and sex

## Appendix 16- Supplementary table 2 for Article 2

Odds ratios of RCC for plasma concentrations of vitamins B2, B6, folate, B12, and homocysteine and methionine, by comparing RCC cases to control group 2 only, *i)* with and *ii)* without adjustment for risk factors, by *iii)* comparing cases to matched controls in a conditional analysis where missing covariate data were excluded in a sensitivity analysis, and by *iv)* comparing cases with all controls, adjusting for risk factors.

Quartile (Range)	Odds Ratio (95% Confidence Interval)			
	<i>i)</i> Cases compared to control group 2, minimally adjusted <sup>a</sup> (n = 444/553) <sup>a</sup>	<i>ii)</i> Cases compared to control group 2, adjusted for all risk factor <sup>b</sup> (n = 444/553) <sup>b</sup>	<i>iii)</i> Cases compared to matched controls, adjusted for all risk factors (all missing data excluded) <sup>c</sup> (n = 383/383) <sup>c</sup>	<i>iv)</i> Cases compared to all controls, adjusted for all risk factors <sup>d</sup> (n = 556/1109) <sup>d</sup>
<b>Vitamin B2 (Riboflavin) [nmol/L]<sup>e</sup></b>				
1 (2.57-9.83)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (9.84-14.3)	0.72 (0.50-1.02)	0.76 (0.53 - 1.11)	0.78 (0.51 - 1.19)	0.74 (0.55 - 1.00)
3 (14.4-22.2)	0.81 (0.55-1.19)	0.84 (0.56 - 1.25)	0.71 (0.46 - 1.09)	0.82 (0.60 - 1.12)
4 (22.2-416)	0.78 (0.53-1.15)	0.87 (0.58 - 1.31)	0.79 (0.51 - 1.24)	0.78 (0.56 - 1.07)
<i>P for trend<sup>f</sup></i>	0.47	0.96	0.19	0.31
<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>e</sup></b>				
1 (5.95-25.4)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (25.4-35.9)	0.64 (0.46-0.91)	0.69 (0.47 - 0.99)	0.67 (0.44 - 1.02)	0.65 (0.49 - 0.88)
3 (35.9-51.9)	0.60 (0.42-0.86)	0.63 (0.43 - 0.93)	0.58 (0.37 - 0.90)	0.62 (0.46 - 0.84)
4 (51.9-436)	0.52 (0.34-0.77)	0.55 (0.35 - 0.84)	0.34 (0.21 - 0.57)	0.49 (0.35 - 0.68)
<i>P for trend<sup>f</sup></i>	0.002	0.008	<0.001	<0.001
<b>Vitamin B9 (Folate) [nmol/L]<sup>e</sup></b>				
1 (0.20-8.40)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (8.41-11.8)	0.72 (0.48-1.08)	0.67 (0.44 - 1.03)	0.93 (0.58 - 1.49)	0.86 (0.62 - 1.17)
3 (11.9-17.3)	0.52 (0.35-0.77)	0.52 (0.34 - 0.79)	0.95 (0.58 - 1.55)	0.73 (0.53 - 1.00)
4 (17.3-109)	0.55 (0.37-0.83)	0.54 (0.35 - 0.84)	0.79 (0.47 - 1.31)	0.72 (0.51 - 1.00)
<i>P for trend<sup>f</sup></i>	0.01	0.02	0.07	0.03
<b>Vitamin B12 (Cobalamin) [pmol/L]<sup>e</sup></b>				
1 (75.2-281)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (281-343)	0.89 (0.63-1.26)	0.95 (0.66 - 1.37)	0.94 (0.60 - 1.45)	0.94 (0.70 - 1.27)
3 (344-419)	0.93 (0.64-1.35)	1.03 (0.70 - 1.52)	1.06 (0.68 - 1.64)	0.93 (0.68 - 1.27)
4 (419-5000)	0.99 (0.68-1.44)	1.07 (0.72 - 1.61)	0.76 (0.48 - 1.19)	0.83 (0.60 - 1.13)
<i>P for trend<sup>f</sup></i>	0.5	0.21	0.47	0.61
<b>Methionine [μmol/L]<sup>e</sup></b>				
1 (2.1-21.48)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (21.5-25)	0.84 (0.59-1.21)	0.82 (0.56 - 1.21)	1.13 (0.74 - 1.71)	0.95 (0.70 - 1.28)
3 (25.0-28.9)	0.71 (0.47-1.05)	0.64 (0.42 - 0.98)	0.75 (0.47 - 1.19)	0.71 (0.51 - 0.98)
4 (29.0-71.4)	0.72 (0.49-1.06)	0.74 (0.49 - 1.11)	0.66 (0.40 - 1.07)	0.78 (0.56 - 1.08)
<i>P for trend<sup>f</sup></i>	0.02	0.06	0.07	0.03
<b>Homocysteine [μmol/L]<sup>e</sup></b>				
1 (3.68-8.03)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (8.04-9.87)	0.79 (0.55-1.15)	0.73 (0.50 - 1.08)	1.13 (0.74 - 1.72)	0.88 (0.65 - 1.20)
3 (9.88-11.9)	0.74 (0.49-1.13)	0.67 (0.43 - 1.04)	0.93 (0.59 - 1.48)	0.74 (0.53 - 1.04)
4 (12.0-64.9)	0.91 (0.61-1.37)	0.84 (0.54 - 1.30)	0.84 (0.51 - 1.39)	0.88 (0.63 - 1.24)
<i>P for trend<sup>f</sup></i>	0.70	0.89	0.44	0.76

a) Assessed by analyzing RCC cases and additional unmatched controls by unconditional logistic regression, adjusting for country, sex, and age at recruitment (in 5 year groups). Danish subjects did not contribute to this analysis.

b) Assessed by analyzing RCC cases and additional unmatched controls by unconditional logistic regression after multiple imputation of missing covariate data, adjusting for country, sex, age at recruitment (in 5 year groups), waist-to-hip ratio (quartiles defined among matched controls), hypertension (yes/no), educational attainment (4 categories), smoking status (never/former/current), plasma cotinine (quartiles defined by the distribution for current smokers), alcohol intake at recruitment (g/day) and alcohol intake (ever/never). Danish subjects did not contribute to this analysis.

c) Sensitivity analysis in which participants with missing covariate data were excluded. Assessed by analyzing RCC cases and their individually matched controls by conditional logistic regression, conditioning on individual case set, and adjusting for waist-to-hip ratio (quartiles defined among matched controls), hypertension (yes/no), educational attainment (4 categories), smoking status (never/former/current), plasma cotinine (quartiles defined by the distribution for current smokers), alcohol intake at recruitment (g/day) and alcohol intake (ever/never). Case-control numbers only include those case-sets where both the case and matched control had complete plasma measurements and covariate information.

d) Assessed by analysing RCC cases and all controls combined by unconditional logistic regression after multiple imputation of missing covariate data, adjusting for country, sex, age at recruitment (in 5 year groups), waist-to-hip ratio (quartiles defined among matched controls), hypertension (yes/no), educational attainment (4 categories), smoking status (never/former/current), plasma cotinine (quartiles defined by the distribution for current smokers), alcohol intake at recruitment (g/day) and alcohol intake (ever/never).

e) Quartile cut-off points were determined based on the plasma level distribution of each biomarker for 556 individually matched controls.

f) *P* for trend assessed by the base 2 logarithm of plasma concentrations.



**Appendix 17-** Supplementary table 3 for Article 2

Hazard ratios of all-cause mortality for RCC cases for quartiles of plasma vitamins B2, B6, folate, B12, and homocysteine and methionine: Sensitivity analysis excluding missing data

Quartile (Range)	Deceased <sup>b</sup>	Alive <sup>b</sup>	Person years	Hazard ratio (95% CI)	
				Sensitivity analysis (missing covariate data excluded) Adjusted for risk factors <sup>a</sup>	
<b>Vitamin B2 (Riboflavin) [nmol/L]<sup>c</sup></b>					
1 (2.57-9.83)	50 (24%)	111 (32%)	744.1	1.00 (reference)	
2 (9.84-14.34)	52 (25%)	75 (22%)	602.8	1.35 (0.83 - 2.20)	
3 (14.35-22.16)	48 (23%)	85 (25%)	528.9	1.34 (0.83 - 2.19)	
4 (22.19-416.79)	55 (27%)	73 (21%)	501.4	1.65 (1.00 - 2.74)	
	<i>P for trend<sup>d</sup></i>			0.18	
<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>c</sup></b>					
1 (5.95-25.37)	90 (44%)	116 (34%)	890.4	1.00 (reference)	
2 (25.43-35.9)	51 (25%)	83 (24%)	576.5	0.85 (0.56 - 1.30)	
3 (35.92-51.88)	37 (18%)	84 (24%)	530.8	0.74 (0.46 - 1.19)	
4 (51.92-436.13)	27 (13%)	61 (18%)	379.5	0.54 (0.30 - 0.97)	
	<i>P for trend<sup>d</sup></i>			0.01	
<b>Vitamin B9 (Folate) [nmol/L]<sup>c</sup></b>					
1 (0.2-8.4)	54 (26%)	97 (28%)	697.0	1.00 (reference)	
2 (8.41-11.84)	63 (31%)	74 (22%)	501.5	1.72 (1.07 - 2.76)	
3 (11.86-17.25)	46 (22%)	85 (25%)	594.9	1.06 (0.64 - 1.76)	
4 (17.27-109.35)	42 (20%)	87 (25%)	580.5	0.93 (0.54 - 1.59)	
	<i>P for trend<sup>d</sup></i>			0.53	
<b>Vitamin B12 (Cobalamin) [pmol/L]<sup>c</sup></b>					
1 (75.16-281.28)	61 (30%)	90 (26%)	638.9	1.00 (reference)	
2 (281.37-343.39)	57 (28%)	90 (26%)	600.5	0.70 (0.44 - 1.11)	
3 (343.51-419)	46 (22%)	88 (26%)	595.5	0.64 (0.40 - 1.04)	
4 (419.35-5000)	41 (20%)	75 (22%)	539.1	0.77 (0.47 - 1.27)	
	<i>P for trend<sup>d</sup></i>			0.26	
<b>Methionine [μmol/L]<sup>c</sup></b>					
1 (2.1-21.48)	63 (31%)	82 (24%)	657.7	1.00 (reference)	
2 (21.5-25)	49 (24%)	115 (33%)	765.7	0.85 (0.54 - 1.35)	
3 (25.01-28.94)	43 (21%)	71 (21%)	455.3	1.06 (0.64 - 1.73)	
4 (28.95-71.4)	50 (24%)	76 (22%)	498.4	1.06 (0.66 - 1.70)	
	<i>P for trend<sup>d</sup></i>			0.6	
<b>Homocysteine [μmol/L]<sup>c</sup></b>					
1 (3.68-8.03)	36 (18%)	103 (30%)	658.3	1.00 (reference)	
2 (8.04-9.87)	58 (28%)	96 (28%)	677.6	1.95 (1.17 - 3.27)	
3 (9.88-11.94)	46 (22%)	68 (20%)	492.5	2.08 (1.23 - 3.53)	
4 (11.95-64.88)	65 (32%)	77 (22%)	548.8	1.67 (0.96 - 2.90)	
	<i>P for trend<sup>d</sup></i>			0.17	

a) Sensitivity analysis in which participants with missing covariate data were excluded. Assessed by analyzing RCC cases by Cox-proportional hazards regression, adjusting for country, sex, and age at diagnosis, and further by hypertension (yes/no), waist-to-hip ratio (quartiles), educational attainment (4 categories), smoking status (never/former/current), plasma cotinine (quartiles defined by the distribution for current smokers), alcohol intake at recruitment (g/day), and body mass index (three categories were defined: <25 kg/m<sup>2</sup>; 25-30 kg/m<sup>2</sup>; >30 kg/m<sup>2</sup>, in order to have a reasonable distribution of body mass index in each group).

b) Vital status for RCC case at the last follow-up.

c) Quartile cut-off points were determined based on the plasma level distribution of each biomarker for 556 individually matched controls.

d) P for trend assessed by the base 2 logarithm of plasma concentrations.

## Appendix 18- Supplementary table 4 for Article 2

Distribution of Pyridoxal 5'-phosphate demographic, baseline, and tumour characteristics

	Pyridoxal 5'-phosphate [nmol/L]			
	RCC cases (n = 556)		Matched controls (n = 556)	
	No. of cases	Median (5th to 95th percentile)	No. of controls	Median (5th to 95th percentile)
<b>Gender</b>				
Men	310	31.9 (13.1-83.1)	310	38.7 (16.6-110)
Women	246	28.8 (12.1-116.6)	246	32.4 (13.6-122.1)
<b>Participating countries</b>				
France	13	32.2 (11.2-314.8)	13	31.4 (23.2-128.3)
Italy	88	25.1 (11.9-58.2)	88	31.7 (12.4-69.8)
Spain	52	29.5 (17.3-53.3)	52	37.4 (19.2-71.4)
United Kingdom	67	35.5 (11.9-123.7)	67	38.6 (16.8-203.9)
The Netherlands	46	28.0 (14.2-77.4)	46	26.5 (14.2-91.3)
Greece	17	23.1 (9.0-64.9)	17	28.5 (13.3-245)
Germany	125	34.0 (15.0-104)	125	35.6 (13.9-78.7)
Sweden	32	30.9 (15.4-148.1)	32	43.2 (19.5-270.1)
Denmark	112	30.6 (13-101.1)	112	46.6 (17.6-194.6)
Norway	4	27.1 (9.9-39.2)	4	36.4 (19.8-51.6)
<b>Smoking status</b>				
Never smokers	225	31.9 (13.9-104)	244	36.0 (16.2-98.0)
Former smokers	160	33.7 (14.2-84.9)	180	37.7 (16.1-95.7)
Current smokers	166	26.2 (11.7-65.1)	129	33.8 (13.7-150.1)
<b>Educational attainment</b>				
Primary school	229	27.3 (13.2-66.9)	206	33.8 (13.6-127)
Technical/professional school	124	31.7 (11.9-86.8)	136	37.9 (14-122.1)
Secondary school	77	30.8 (12.1-66.4)	66	31.6 (16.7-89.2)
Higher education	110	35.7 (16.2-117)	134	39.0 (17.0-110)
<b>Body mass index</b>				
<18.5	2	25.9 (7.3-44.5)	2	34.9 (18.1-51.7)
18.5-25	179	31.9 (13-148.1)	221	36.0 (16.4-128.3)
25-30	247	32.8 (13.6-81.3)	241	37.0 (16.4-101.5)
30-35	99	27.5 (11.1-80.6)	69	36.4 (16.6-176.1)
≥35	29	26.0 (13.2-45.3)	23	20.4 (9.5-52.7)
<b>Waist to hip ratio</b>				
0.56-0.79	92	32.5 (11.9-89)	113	34.4 (16.2-127)
0.80-0.89	129	31.5 (13.7-116.6)	145	33.1 (13.3-121.9)
0.90-0.94	103	32.7 (13.9-66.2)	114	38.4 (16.6-145.8)
0.95-1.30	196	28.4 (12.9-86.8)	148	38.3 (15.7-101.1)
<b>Hypertension</b>				
No	276	29.8 (13.6-91.6)	325	37.8 (14.2-128.3)
Yes	192	31.0 (12.9-86.8)	140	33.0 (14.1-89.3)
<b>Alcohol intake</b>				
Never drinkers	37	28.4 (10.3-172.5)	22	33.5 (18.9-72.5)
Ever drinkers	509	30.8 (13.6-86.8)	530	36.7 (14.6-121.9)
<b>Alcohol intake at recruitment</b>				
<5 g/day	248	27.9 (11.9-102.6)	221	31.7 (13.5-108.5)
5-20 g/day	156	30.0 (13.0-83.1)	173	37.9 (16.8-121.9)
≥20 g/day	152	34 (14.2-86.8)	162	41.3 (17.5-108.9)
<b>Stage of the tumour</b>				
Localised	161	30.4 (13.9-102.6)		
Metastatic	45	30.5 (14.7-66.2)		
Metastatic regional	18	33.2 (12.9-183.9)		
Metastatic distant	27	27.9 (13.2-81.3)		
Unknown	305	30.7 (13.0-83.5)		

## Appendix 19- Supplementary table 5 for Article 2

## Distribution of staging by selected variables

	<b>Stage Available</b>	
	<b>Yes (n=251)</b>	<b>No (n=305)</b>
<b>Gender</b>		
Men	153 (61%)	157 (51.5%)
Women	98 (39%)	148 (48.5%)
<b>Participating countries</b>		
France	1 (0.4%)	12 (3.9%)
Italy	33 (13.1%)	55 (18%)
Spain	37 (14.7%)	15 (4.9%)
United Kingdom	21 (8.4%)	46 (15.1%)
The Netherlands	0 (0%)	46 (15.1%)
Greece	12 (4.8%)	5 (1.6%)
Germany	95 (37.8%)	30 (9.8%)
Sweden	0 (0%)	32 (10.5%)
Denmark	49 (19.5%)	63 (20.7%)
Norway	3 (1.2%)	1 (0.3%)
<b>Smoking status</b>		
Never smokers	100 (39.8%)	125 (41%)
Former smokers	68 (27.1%)	92 (30.2%)
Current smokers	81 (32.3%)	85 (27.9%)
Unknown	2 (0.8%)	3 (1%)
<b>Educational attainment</b>		
Primary school	116 (46.2%)	113 (37%)
Technical/professional school	61 (24.3%)	63 (20.7%)
Secondary school	24 (9.6%)	53 (17.4%)
Higher education	48 (19.1%)	62 (20.3%)
Unknown	2 (0.8%)	14 (4.6%)
<b>Body mass index</b>		
<18.5	1 (0.4%)	1 (0.3%)
18.5-25	68 (27.1%)	111 (36.4%)
25-30	117 (46.6%)	130 (42.6%)
30-35	50 (19.9%)	49 (16.1%)
≥35	15 (6%)	14 (4.6%)
<b>Waist to hip ratio</b>		
0.56-0.79	32 (12.7%)	60 (19.7%)
0.80-0.89	58 (23.1%)	71 (23.3%)
0.90-0.94	56 (22.3%)	47 (15.4%)
0.95-1.30	102 (40.6%)	94 (30.8%)
Unknown	3 (1.2%)	33 (10.8%)
<b>Hypertension</b>		
No	132 (52.6%)	144 (47.2%)
Yes	94 (37.5%)	98 (32.1%)
Unknown	25 (10%)	63 (20.7%)
<b>Alcohol intake</b>		
Never drinkers	20 (8%)	17 (5.6%)
Ever drinkers	230 (91.6%)	279 (91.5%)
Unknown	1 (0.4%)	9 (3%)
<b>Alcohol intake at recruitment</b>		
<5 g/day	101 (40.2%)	147 (48.2%)
5-20 g/day	71 (28.3%)	85 (27.9%)
≥20 g/day	79 (31.5%)	73 (23.9%)

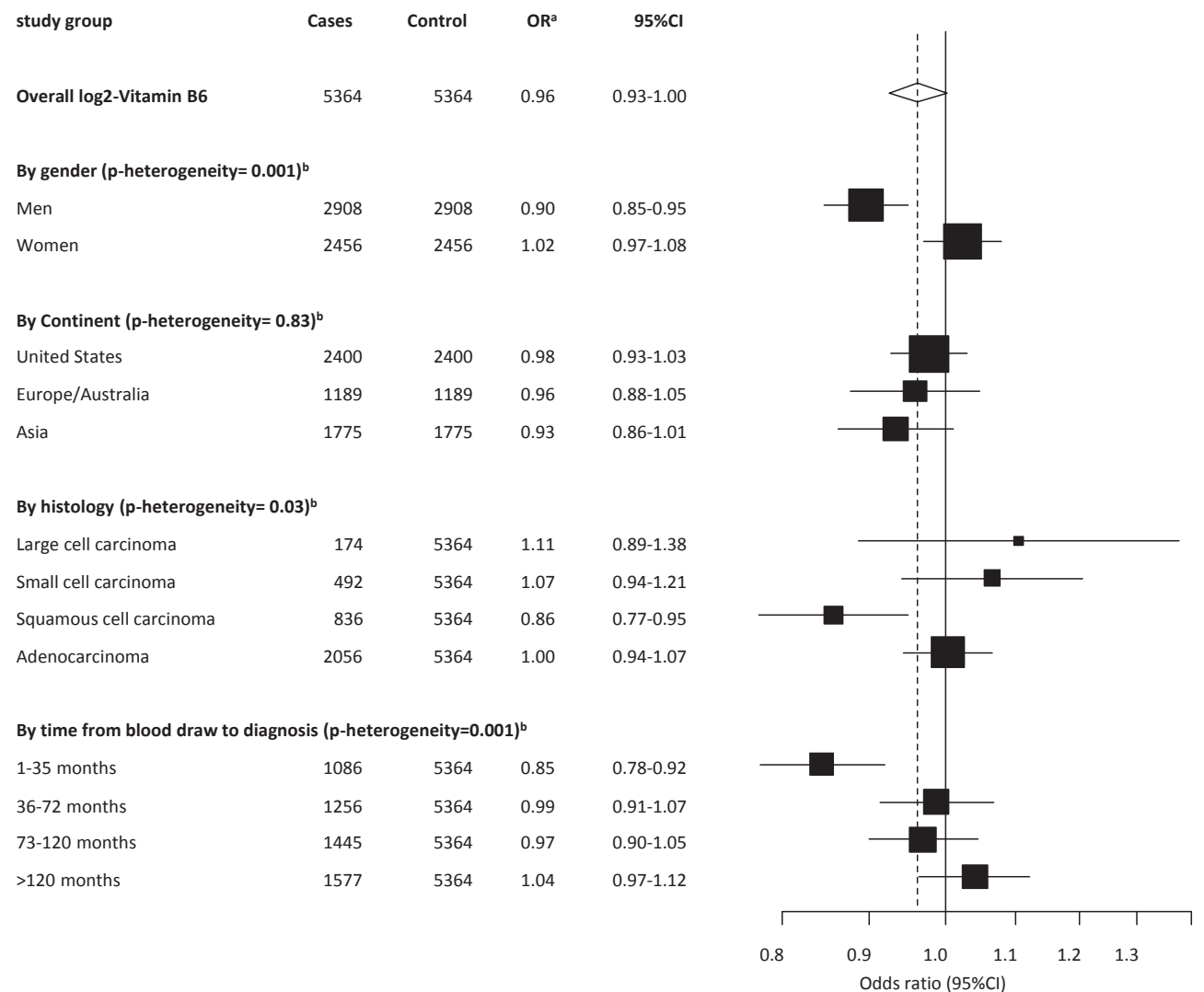
**Age at recruitment**

< 40	6 (2.4%)	7 (2.3%)
40-45	19 (7.6%)	16 (5.2%)
45-50	20 (8%)	27 (8.9%)
50-55	53 (21.1%)	79 (25.9%)
55-60	55 (21.9%)	66 (21.6%)
65-70	81 (32.3%)	81 (26.6%)
75-80	11 (4.4%)	21 (6.9%)
> 80	6 (2.4%)	8 (2.6%)

---

**Appendix 20-** Supplementary figure 1 for Article 3

Forest plot showing stratified ORs of lung cancer for log<sub>2</sub>- of circulating vitamin B6 overall.

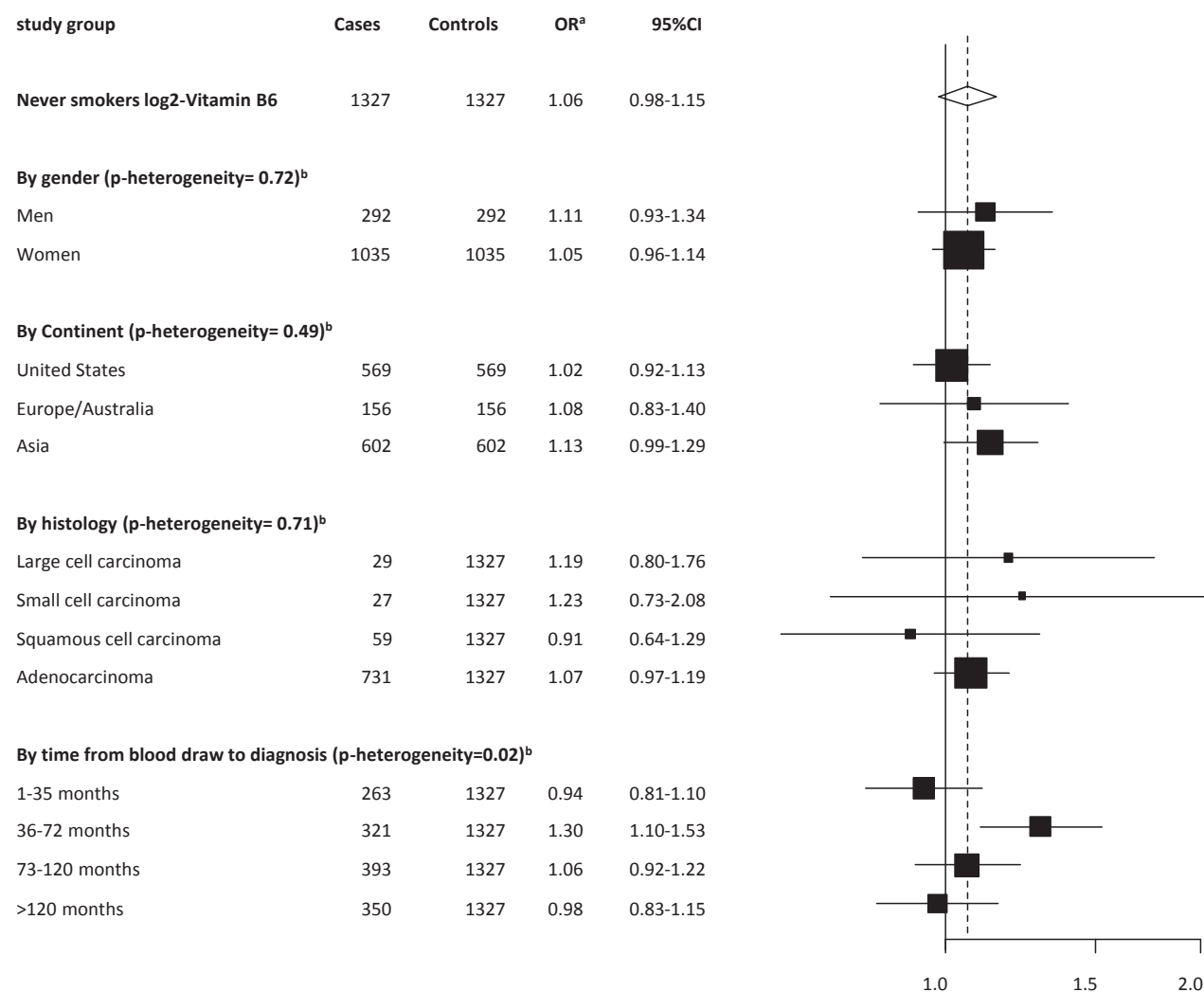


<sup>a</sup>ORs were assessed by conditional logistic regression by including the base 2 logarithm of circulating concentrations (ORs indicate relative risks of a doubling in circulating concentrations), and where relevant adjusted for cotinine and education; the black squares and horizontal lines indicate the ORs and 95% CIs; the size of the black squares is proportional to the inverse variance of the logistic regression estimates.

<sup>b</sup>P<sub>heterogeneity</sub> indicates results of chi-square test assessing the null hypothesis of ORs being identical.

**Appendix 21-** Supplementary figure 2 for Article 3

Forest plot showing stratified ORs of lung cancer for log<sub>2</sub>- of circulating vitamin B6 among never smokers.

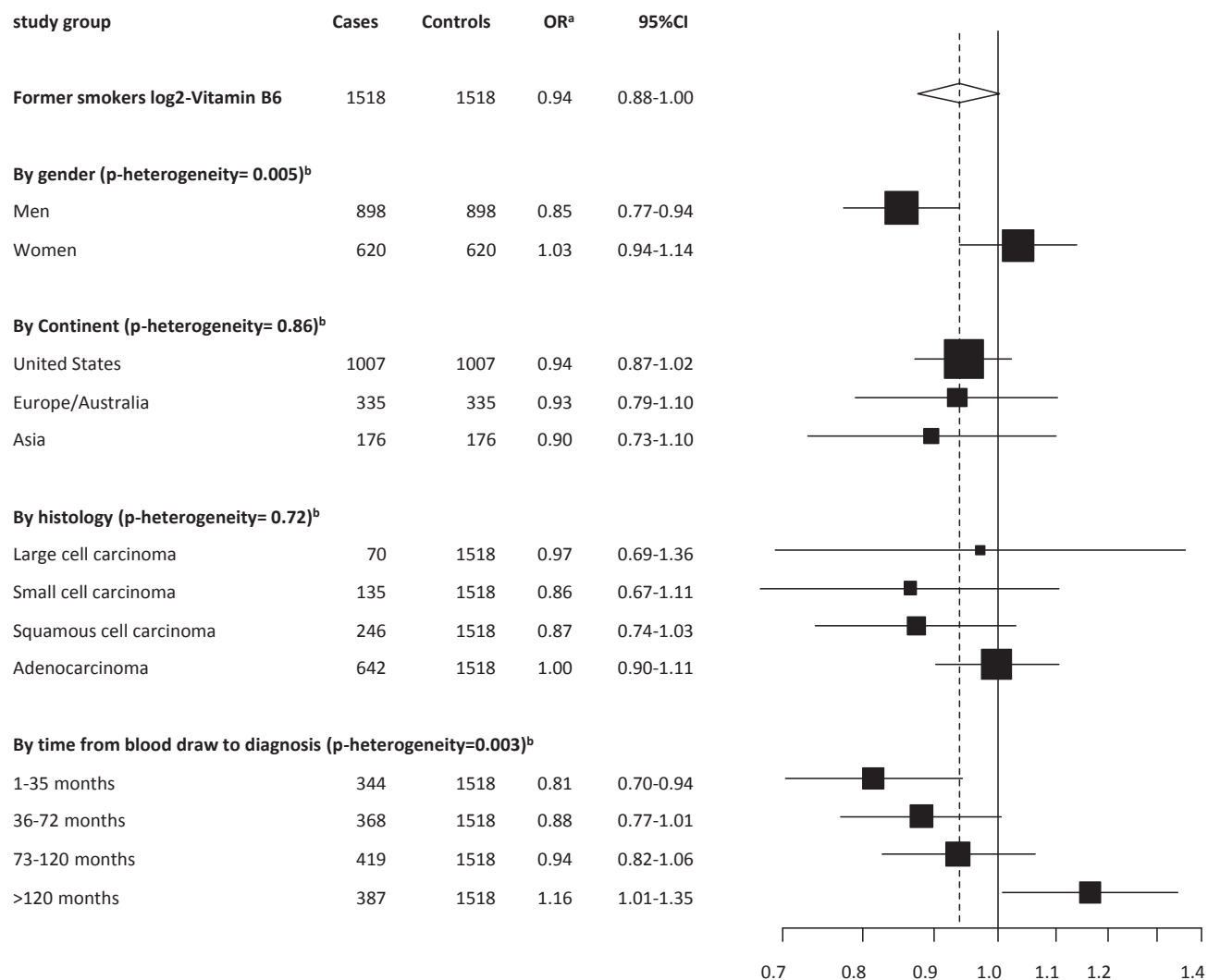


<sup>a</sup>ORs were assessed by conditional logistic regression by including the base 2 logarithm of circulating concentrations (ORs indicate relative risks of a doubling in circulating concentrations), and where relevant adjusted for cotinine and education; the black squares and horizontal lines indicate the ORs and 95% CIs; the size of the black squares is proportional to the inverse variance of the logistic regression estimates.

<sup>b</sup>P<sub>heterogeneity</sub> indicates results of chi-square test assessing the null hypothesis of ORs being identical.

**Appendix 22-** Supplementary figure 3 for Article 3

Forest plot showing stratified ORs of lung cancer for log<sub>2</sub>- of circulating vitamin B6 among former smokers.

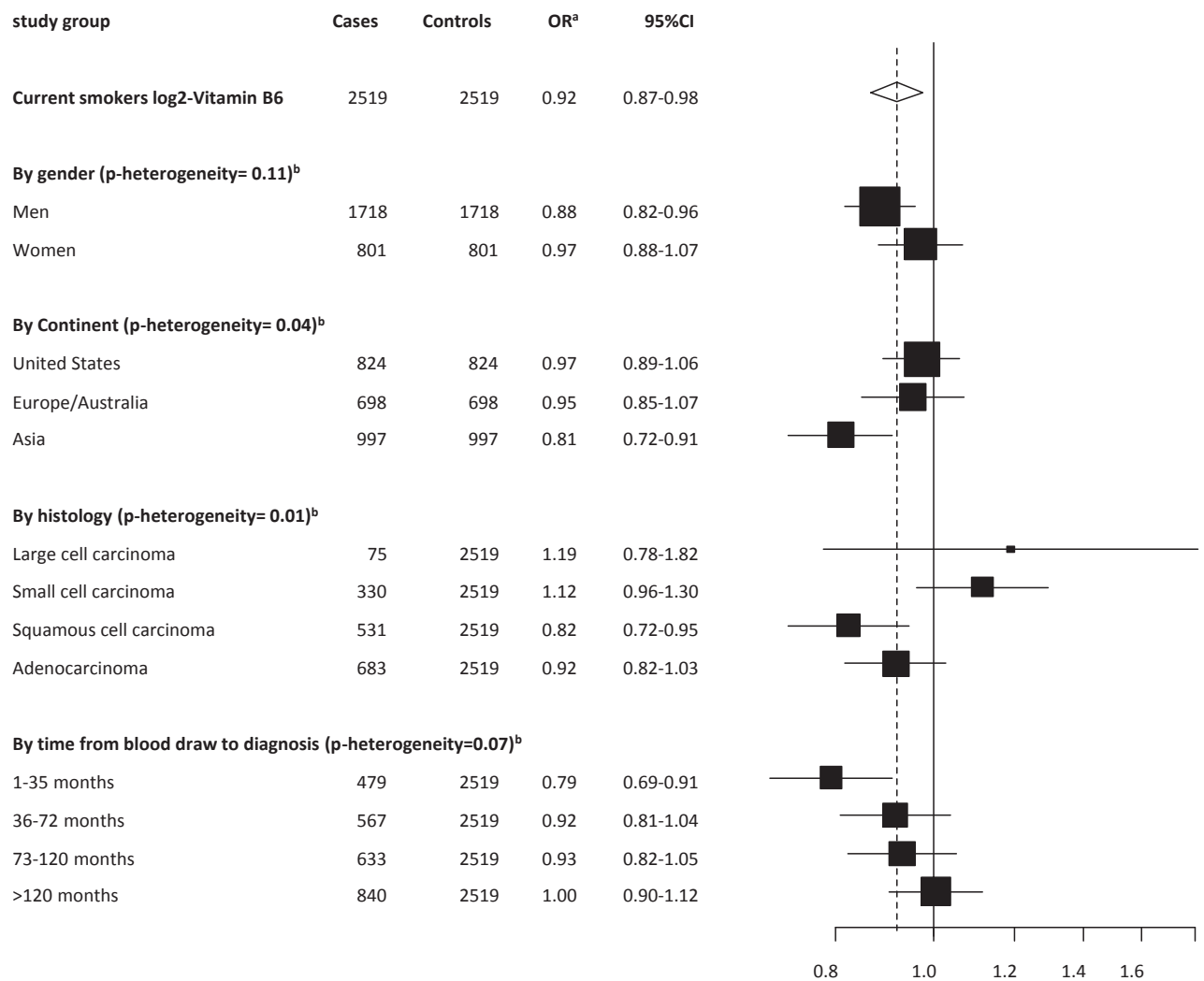


<sup>a</sup>ORs were assessed by conditional logistic regression by including the base 2 logarithm of circulating concentrations (ORs indicate relative risks of a doubling in circulating concentrations), and where relevant adjusted for cotinine and education; the black squares and horizontal lines indicate the ORs and 95% CIs; the size of the black squares is proportional to the inverse variance of the logistic regression estimates.

<sup>b</sup>P<sub>heterogeneity</sub> indicates results of chi-square test assessing the null hypothesis of ORs being identical.

**Appendix 23-** Supplementary figure 4 for Article 3

Forest plot showing stratified ORs of lung cancer for log<sub>2</sub>- of circulating vitamin B6 among current smokers.



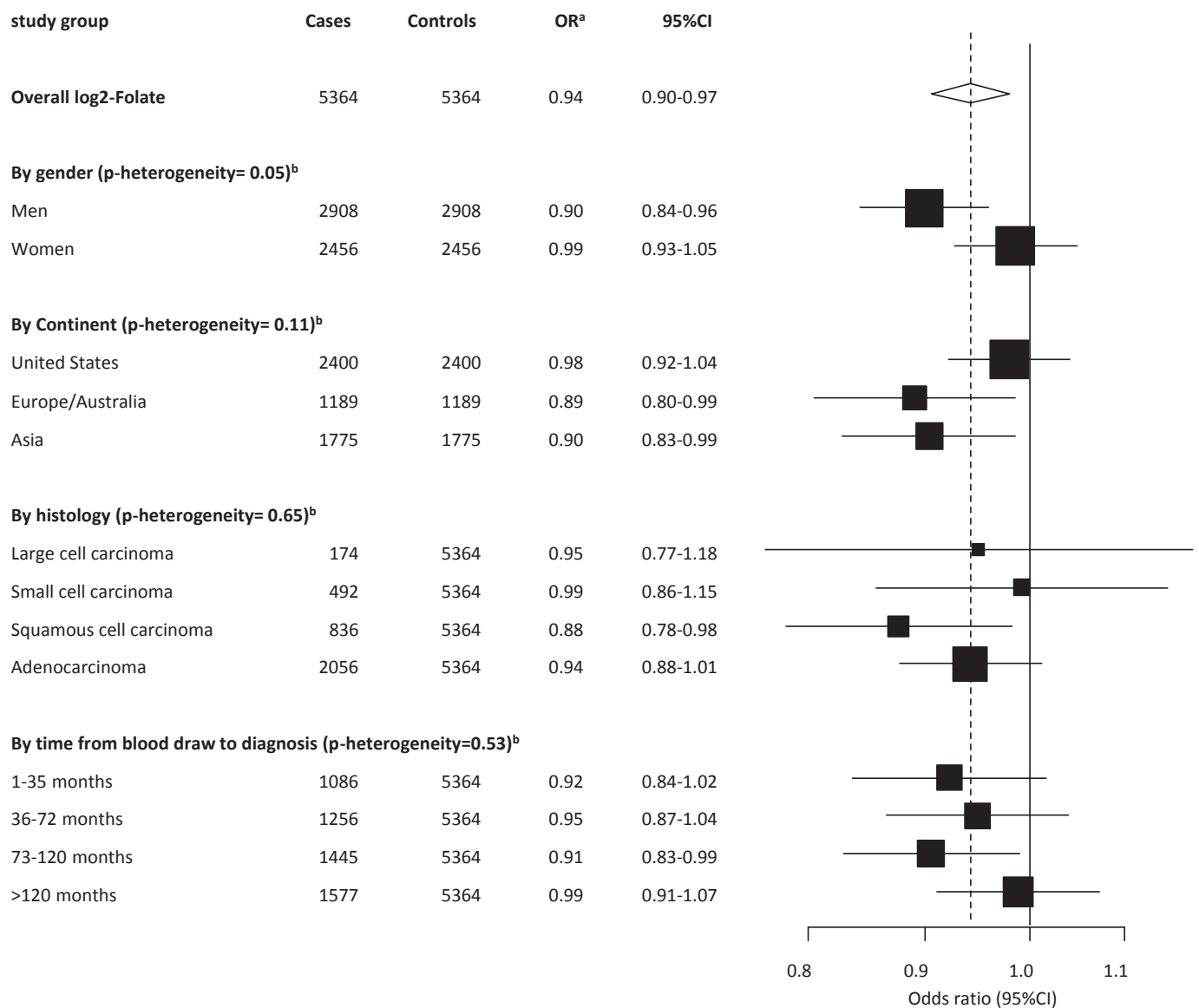
<sup>a</sup>ORs were assessed by conditional logistic regression by including the base 2 logarithm of circulating concentrations (ORs indicate relative risks of a doubling in circulating concentrations), and where relevant adjusted for cotinine and education; the black squares and horizontal lines indicate the ORs and 95% CIs; the size of the black squares is proportional to the inverse variance of the logistic regression estimates.

<sup>b</sup>P<sub>heterogeneity</sub> indicates results of chi-square test assessing the null hypothesis of ORs being identical.



**Appendix 24-** Supplementary figure 5 for Article 3

Forest plot showing stratified ORs of lung cancer for log<sub>2</sub>- of circulating folate overall.

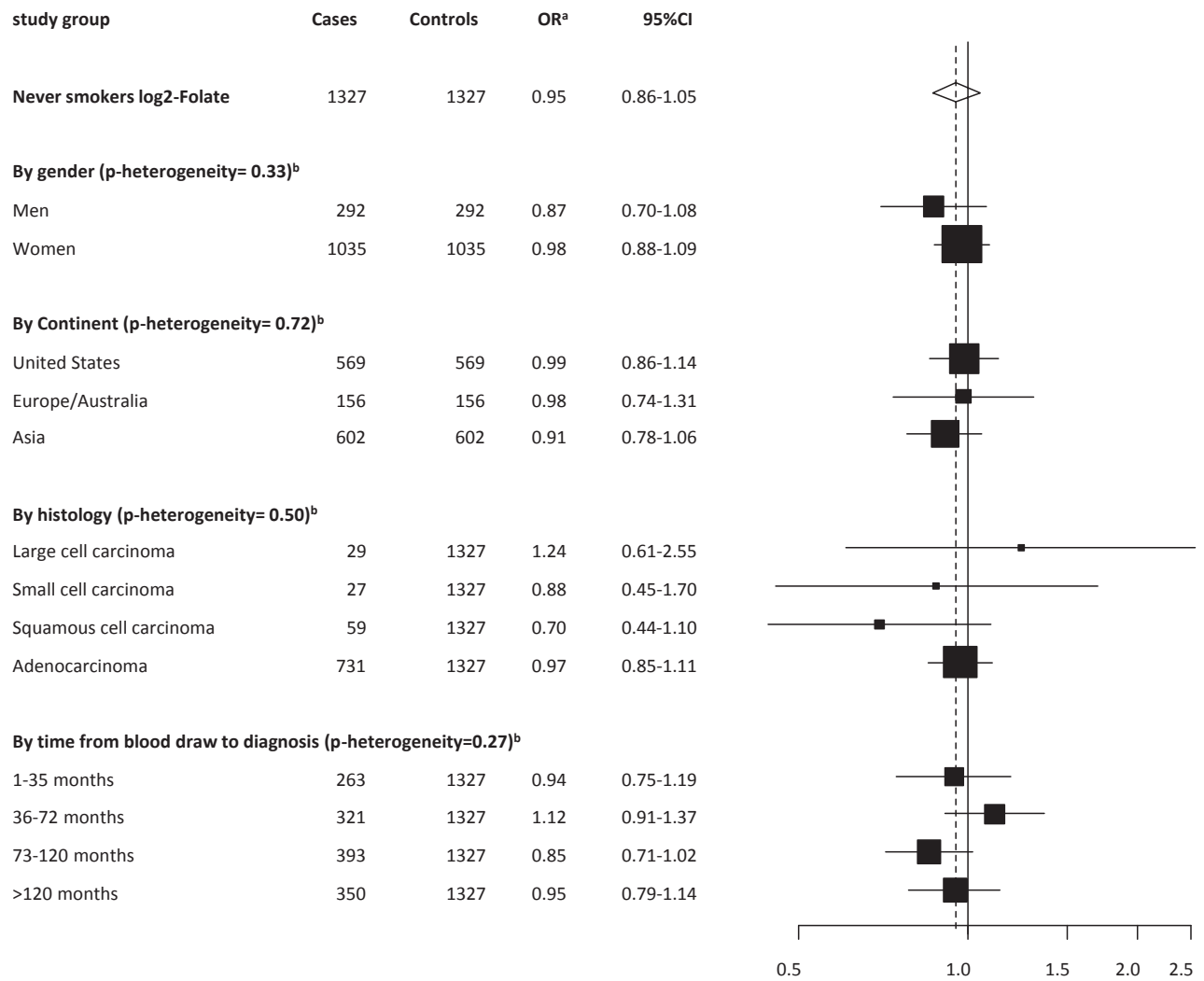


<sup>a</sup>ORs were assessed by conditional logistic regression by including the base 2 logarithm of circulating concentrations (ORs indicate relative risks of a doubling in circulating concentrations), and where relevant adjusted for cotinine and education; the black squares and horizontal lines indicate the ORs and 95% CIs; the size of the black squares is proportional to the inverse variance of the logistic regression estimates.

<sup>b</sup>P<sub>heterogeneity</sub> indicates results of chi-square test assessing the null hypothesis of ORs being identical.

**Appendix 25-** Supplementary figure 6 for Article 3

Forest plot showing stratified ORs of lung cancer for log<sub>2</sub>- of circulating folate among never smokers.

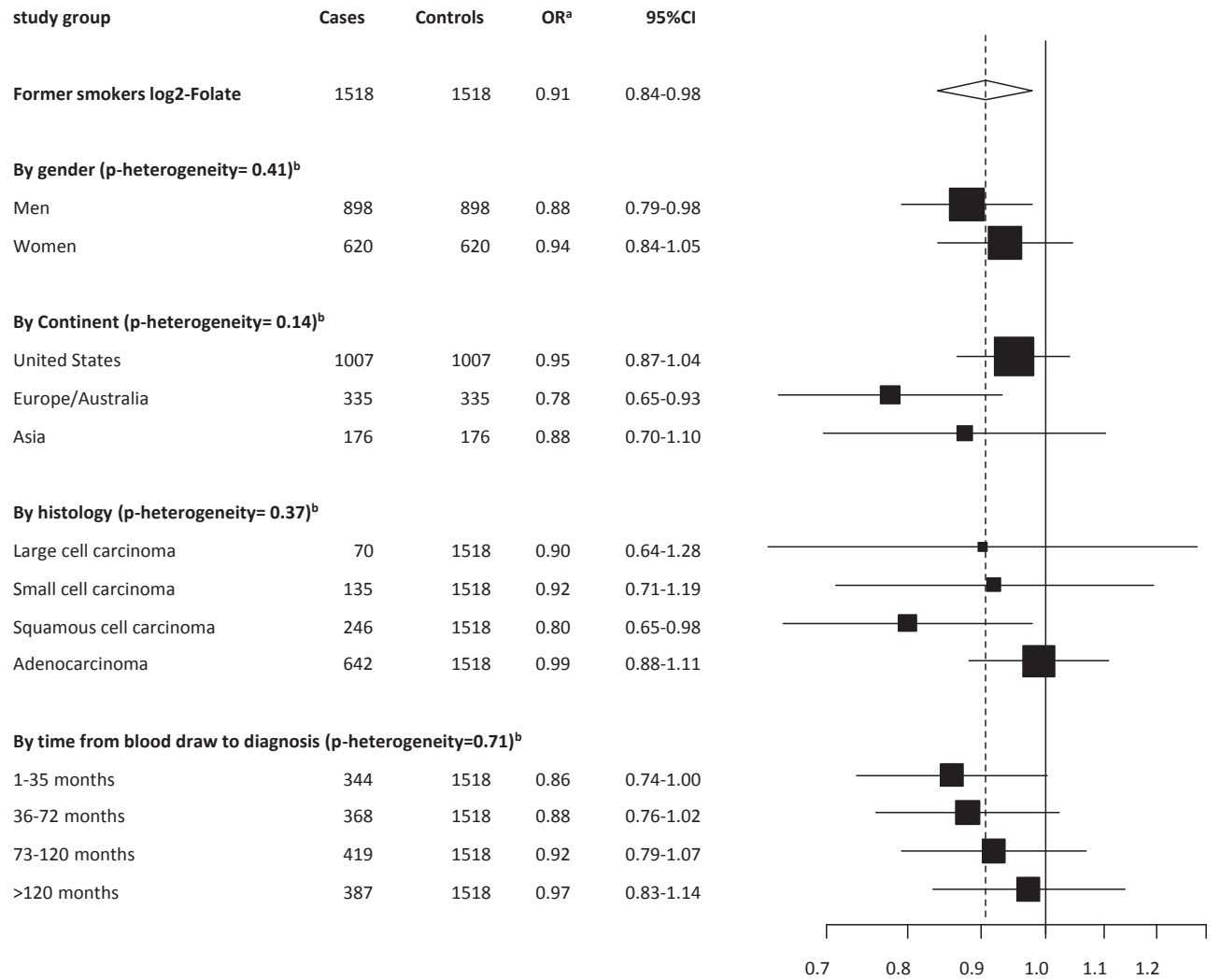


<sup>a</sup>ORs were assessed by conditional logistic regression by including the base 2 logarithm of circulating concentrations (ORs indicate relative risks of a doubling in circulating concentrations), and where relevant adjusted for cotinine and education; the black squares and horizontal lines indicate the ORs and 95% CIs; the size of the black squares is proportional to the inverse variance of the logistic regression estimates.

<sup>b</sup>P<sub>heterogeneity</sub> indicates results of chi-square test assessing the null hypothesis of ORs being identical.

**Appendix 26-** Supplementary figure 7 for Article 3

Forest plot showing stratified ORs of lung cancer for log<sub>2</sub>- of circulating folate among former smokers.

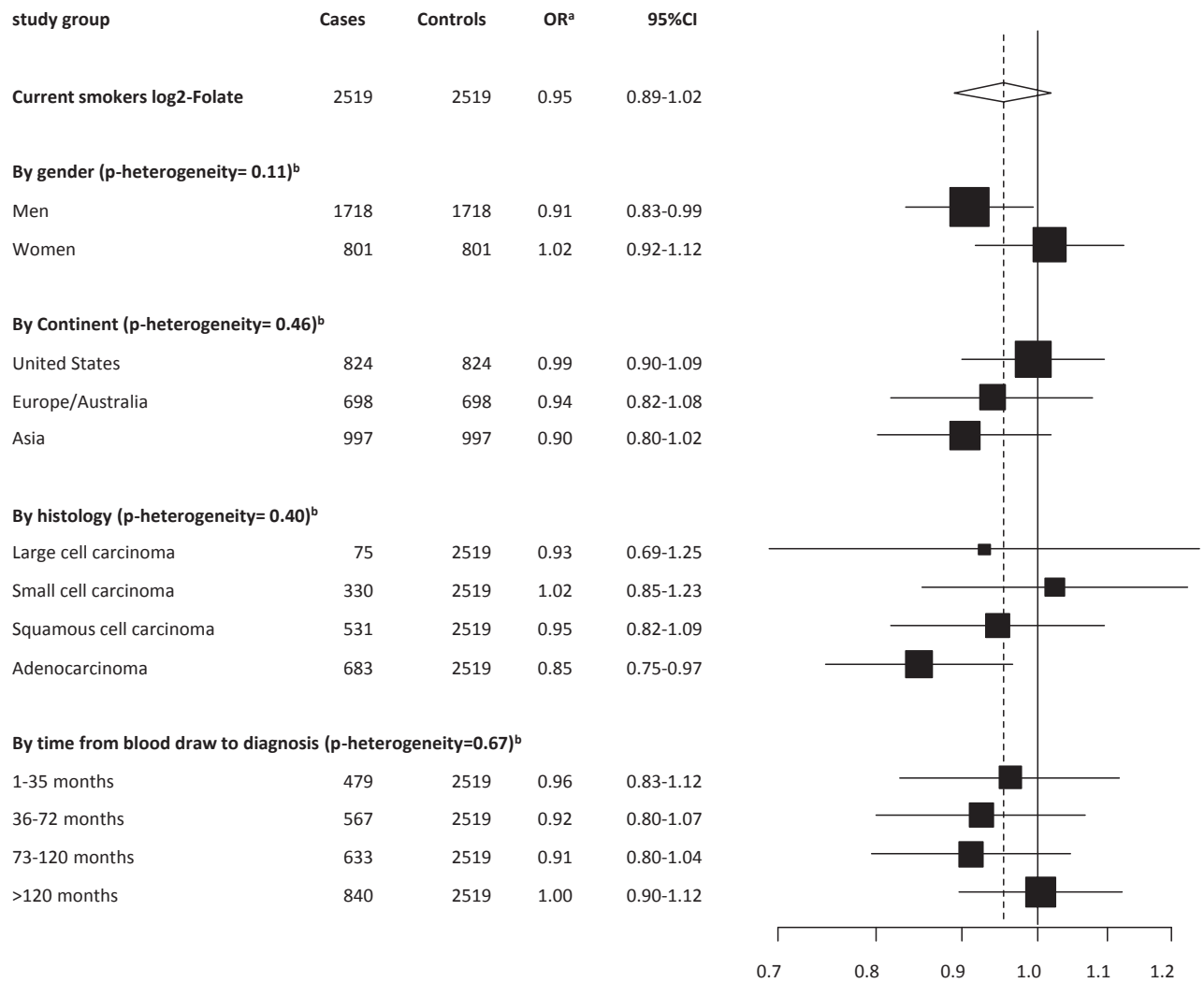


<sup>a</sup>ORs were assessed by conditional logistic regression by including the base 2 logarithm of circulating concentrations (ORs indicate relative risks of a doubling in circulating concentrations), and where relevant adjusted for cotinine and education; the black squares and horizontal lines indicate the ORs and 95% CIs; the size of the black squares is proportional to the inverse variance of the logistic regression estimates.

<sup>b</sup>P<sub>heterogeneity</sub> indicates results of chi-square test assessing the null hypothesis of ORs being identical.

**Appendix 27-** Supplementary figure 8 for Article 3

Forest plot showing stratified ORs of lung cancer for log<sub>2</sub>- of circulating folate among current smokers.



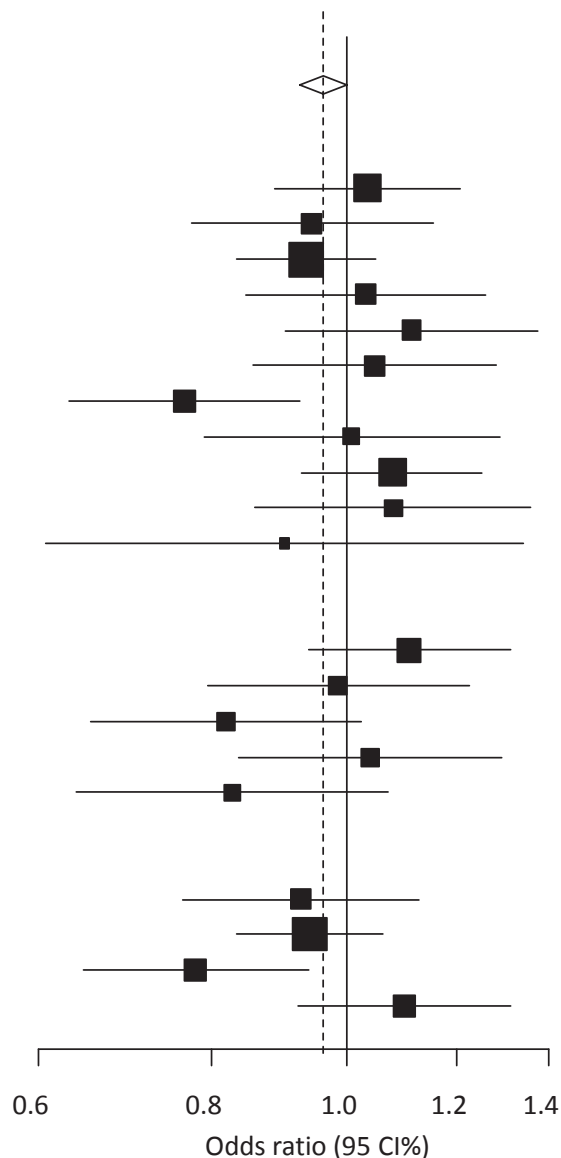
<sup>a</sup>ORs were assessed by conditional logistic regression by including the base 2 logarithm of circulating concentrations (ORs indicate relative risks of a doubling in circulating concentrations), and where relevant adjusted for cotinine and education; the black squares and horizontal lines indicate the ORs and 95% CIs; the size of the black squares is proportional to the inverse variance of the logistic regression estimates.

<sup>b</sup>P<sub>heterogeneity</sub> indicates results of chi-square test assessing the null hypothesis of ORs being identical.

**Appendix 28-** Supplementary figure 9 for Article 3

Forest plot showing stratified ORs of lung cancer for log2 of circulating vitamin B6 by LC3 cohorts.

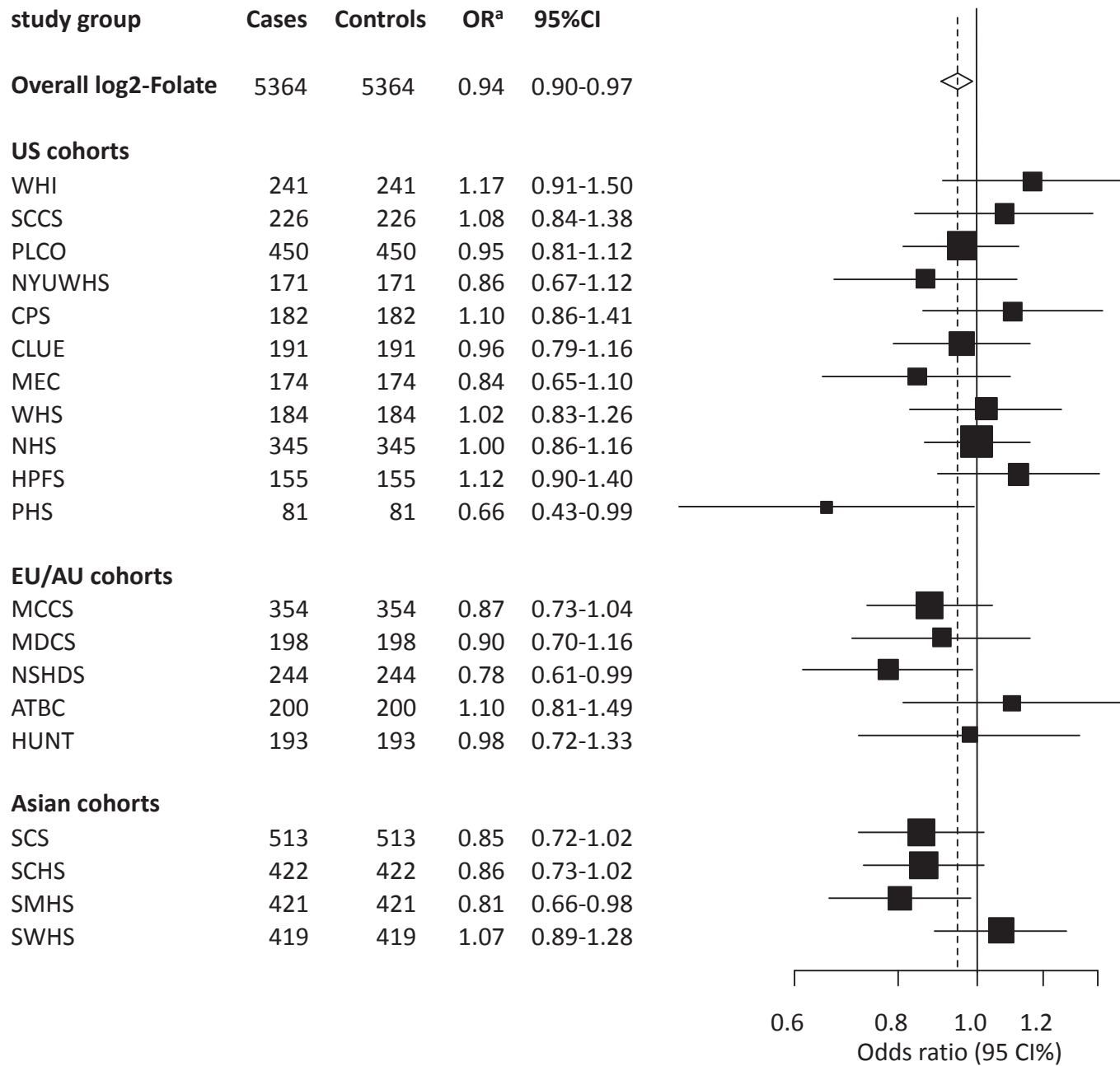
study group	Cases	Controls	OR <sup>a</sup>	95%CI
<b>Overall log2-Vitamin B6</b>	5364	5364	0.96	0.93-1.00
<b>US cohorts</b>				
WHI	241	241	1.04	0.89-1.21
SCCS	226	226	0.95	0.77-1.16
PLCO	450	450	0.94	0.83-1.05
NYUWHS	171	171	1.03	0.85-1.26
CPS	182	182	1.12	0.90-1.38
CLUE	191	191	1.05	0.86-1.28
MEC	174	174	0.77	0.63-0.93
WHS	184	184	1.01	0.79-1.29
NHS	345	345	1.08	0.93-1.25
HPFS	155	155	1.08	0.86-1.36
PHS	81	81	0.90	0.61-1.34
<b>EU/AU cohorts</b>				
MCCS	354	354	1.11	0.94-1.32
MDCS	198	198	0.99	0.80-1.23
NSHDS	244	244	0.82	0.66-1.02
ATBC	200	200	1.04	0.84-1.30
HUNT	193	193	0.83	0.64-1.07
<b>Asian cohorts</b>				
SCS	513	513	0.93	0.76-1.13
SCHS	422	422	0.94	0.83-1.06
SMHS	421	421	0.78	0.65-0.94
SWHS	419	419	1.10	0.92-1.32



<sup>a</sup>ORs were assessed by conditional logistic regression by including the base 2 logarithm of circulating concentrations (ORs indicate relative risks of a doubling in circulating concentrations), and where relevant adjusted for cotinine and education; the black squares and horizontal lines indicate the ORs and 95% CIs; the size of the black squares is proportional to the inverse variance of the logistic regression estimates.  $P_{\text{heterogeneity}}$  indicates results of chi-square test assessing the null hypothesis of ORs being identical.

**Appendix 29-** Supplementary figure 10 for Article 3

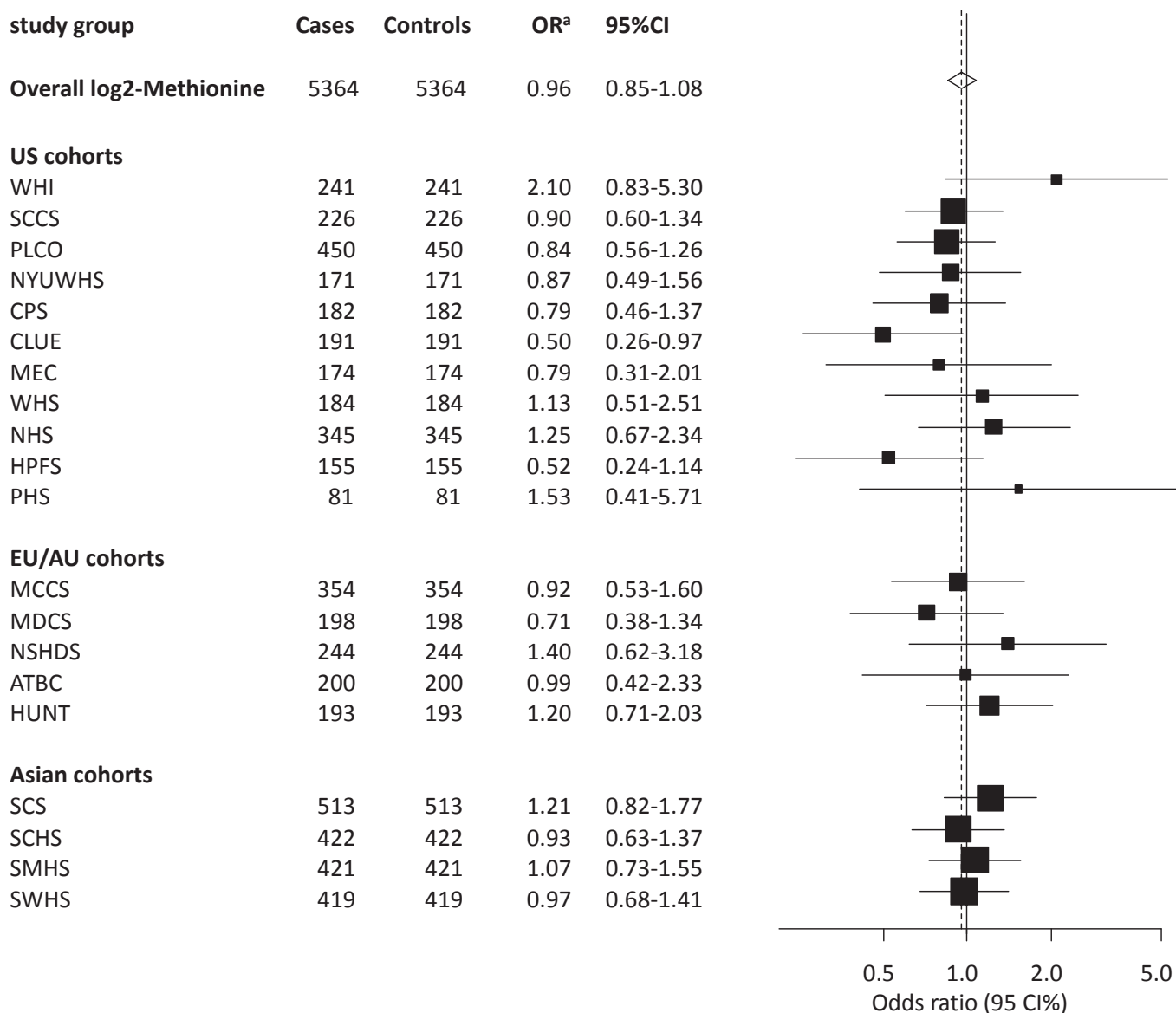
Forest plot showing stratified ORs of lung cancer for log2 of circulating folate by LC3 cohorts.



<sup>a</sup>ORs were assessed by conditional logistic regression by including the base 2 logarithm of circulating concentrations (ORs indicate relative risks of a doubling in circulating concentrations), and where relevant adjusted for cotinine and education; the black squares and horizontal lines indicate the ORs and 95% CIs; the size of the black squares is proportional to the inverse variance of the logistic regression estimates.  $P_{\text{heterogeneity}}$  indicates results of chi-square test assessing the null hypothesis of ORs being identical.

**Appendix 30-** Supplementary figure 11 for Article 3

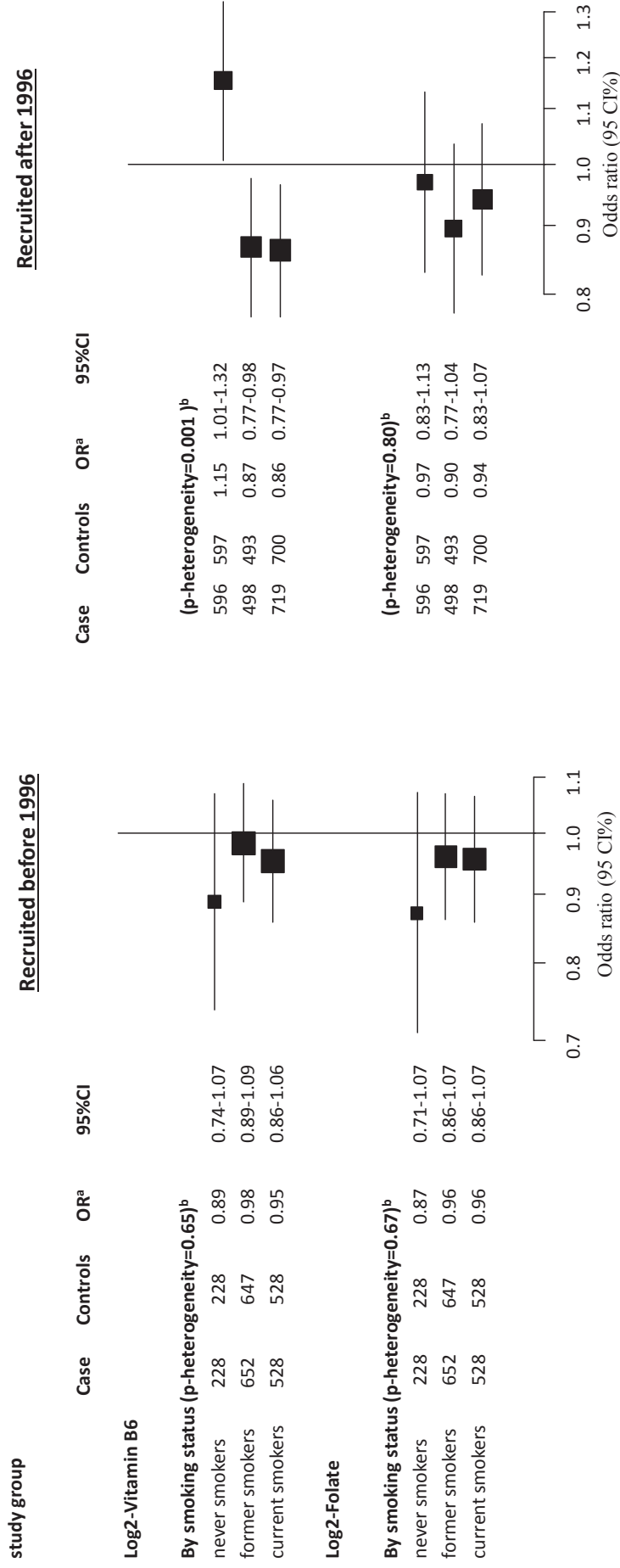
Forest plot showing stratified ORs of lung cancer for log2 of circulating methionine by LC3 cohorts.



<sup>a</sup>ORs were assessed by conditional logistic regression by including the base 2 logarithm of circulating concentrations (ORs indicate relative risks of a doubling in circulating concentrations), and where relevant adjusted for cotinine and education; the black squares and horizontal lines indicate the ORs and 95% CIs; the size of the black squares is proportional to the inverse variance of the logistic regression estimates.  $P_{\text{heterogeneity}}$  indicates results of chi-square test assessing the null hypothesis of ORs being identical.

**Appendix 31-** Supplementary figure 12 for Article 3

Forest plot showing stratified ORs for log2 of circulating vitamin B6 and folate by period of recruitment and by smoking status in US cohorts (n=2400 cases and matched controls).

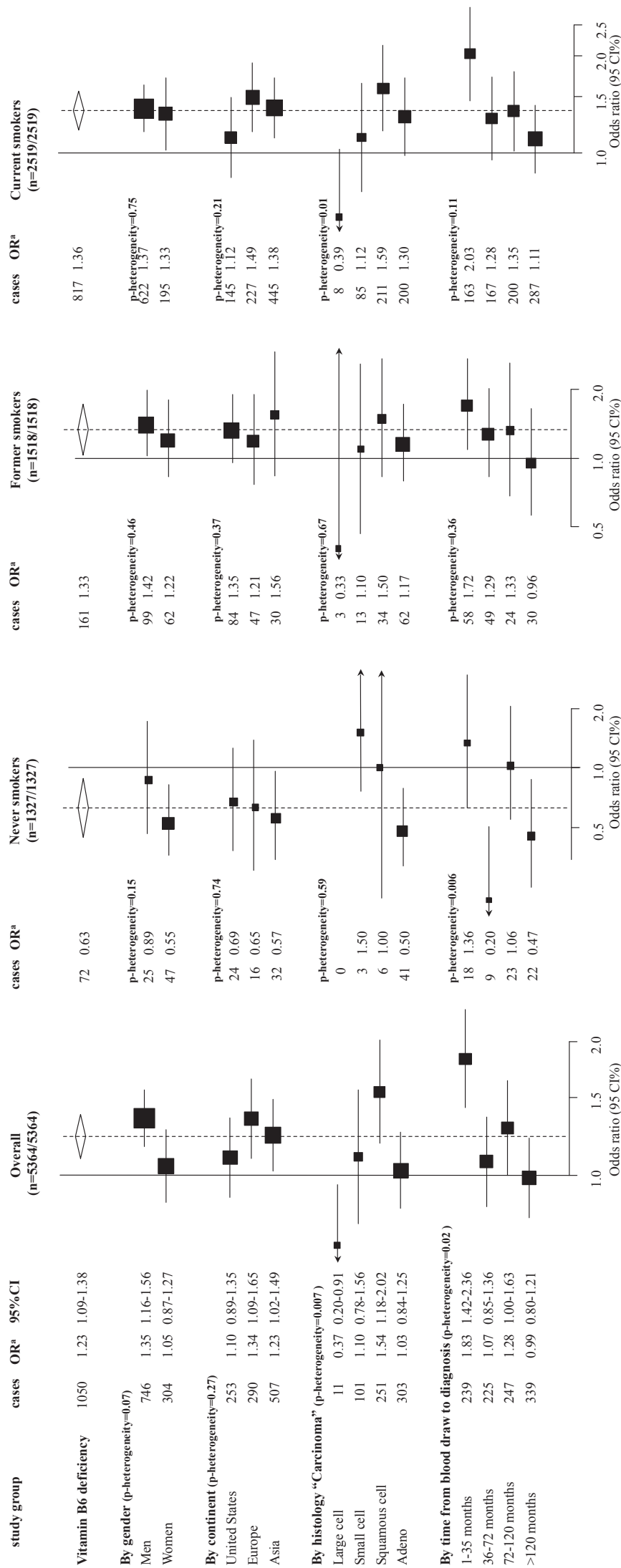


<sup>a</sup>ORs were assessed by conditional logistic regression by including the base 2 logarithm of circulating concentrations (ORs indicate relative risks of a doubling in circulating concentrations), and where relevant adjusted for cotinine and education; the black squares and horizontal lines indicate the ORs and 95% CIs; the size of the black squares are proportional to the inverse variance of logistic regression estimates. <sup>b</sup> P<sub>heterogeneity</sub> indicates results of chi-square test assessing the null hypothesis of ORs being identical.



Appendix 32- Supplementary figure 13 for Article 3

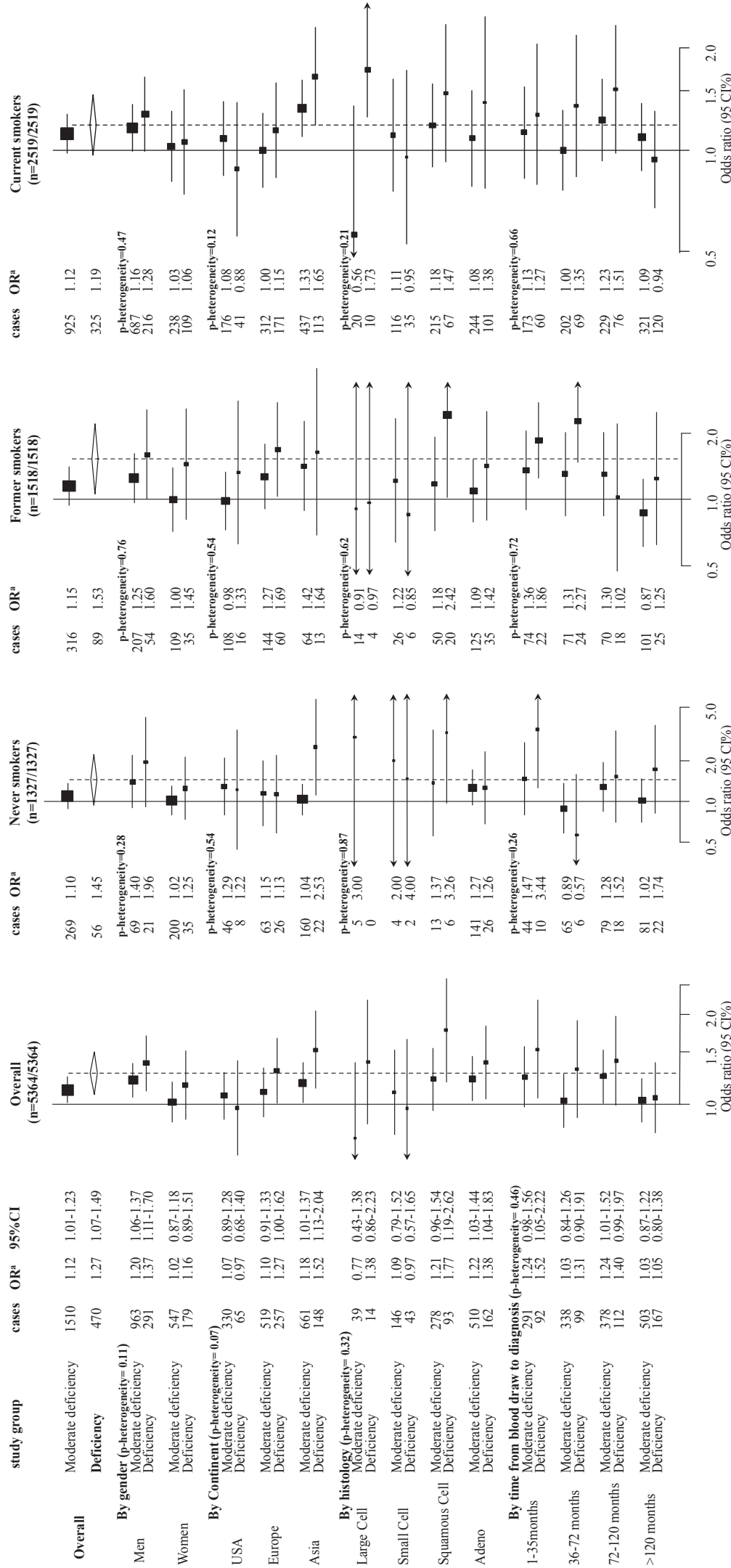
Forest plot showing stratified ORs of lung cancer for B6 deficiency overall and by smoking status.



<sup>a</sup>ORs were assessed by conditional logistic regression, and where relevant adjusted for cotinine and education; the black squares and horizontal lines indicate the ORs and 95% CIs; the size of the black squares are proportional to the inverse variance of logistic regression estimates. Cases listed are those from the group presenting a B6 deficiency. Number of cases within the reference group is not listed. Subjects were considered vitamin B6 deficient if their circulating concentrations were less than 20nmol/L and normal if above or equal to 20nmol/L.

Appendix 33- Supplementary figure 14 for Article 3

Forest plot showing stratified ORs of lung cancer for Folate deficiency overall and by smoking status.



<sup>a</sup>ORs were assessed by conditional logistic regression, and where relevant adjusted for cotinine and education; the black squares and horizontal lines indicate the ORs and 95% CIs; the size of the black squares are proportional to the inverse variance of logistic regression estimates. Cases listed are those from the group presenting a folate deficiency or a moderate deficiency. Number of cases within the reference group is not listed. For folate deficiency, subjects were classified in three categories; deficient (folate <7 nmol/L), moderately deficient (folate 7-13 nmol/L), or normal (folate ≥ 13 nmol/L).

Appendix 34- Supplementary table 1 for Article 3

Odds ratios of lung cancer for circulating levels of vitamin B6, folate and methionine among ever smoking participants in US

		US cohorts				
		Cases	Controls	Model 1 <sup>a</sup> OR (95%CI)	Model 2 <sup>b</sup> OR (95%CI)	Model 3 <sup>c</sup> OR (95%CI)
All ever smokers	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (5.08-29.89)	487	455	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (29.90-49.90)	386	387	0.96 (0.79 - 1.17)	0.98 (0.8 - 1.2)	0.96 (0.78 - 1.18)
	3 (49.91-90.05)	328	343	0.93 (0.76 - 1.14)	0.95 (0.76 - 1.17)	0.91 (0.74 - 1.13)
	4 > 90.05	322	338	0.92 (0.74 - 1.14)	0.97 (0.77 - 1.21)	0.97 (0.78 - 1.22)
	<i>P for trend<sup>e</sup></i>			0.32	0.61	0.57
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (1.70-17.45)	450	429	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (17.46-33.35)	369	388	0.93 (0.76 - 1.14)	0.96 (0.78 - 1.19)	0.95 (0.77 - 1.17)
	3 (33.36-59.10)	353	350	0.98 (0.78 - 1.22)	1.01 (0.8 - 1.27)	1 (0.8 - 1.26)
	4 > 59.10	351	356	0.98 (0.77 - 1.25)	1.03 (0.81 - 1.32)	1.03 (0.81 - 1.32)
	<i>P for trend<sup>e</sup></i>			0.75	0.96	0.97
	<b>Methionine [μmol/L]<sup>d</sup></b>					
	1 (11.83-22.77)	385	363	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.78-26.16)	375	367	0.94 (0.76 - 1.16)	0.95 (0.77 - 1.18)	0.95 (0.77 - 1.18)
3 (26.17-30.47)	372	366	0.96 (0.77 - 1.18)	0.95 (0.76 - 1.19)	0.98 (0.79 - 1.23)	
4 > 30.47	391	427	0.85 (0.69 - 1.05)	0.85 (0.69 - 1.06)	0.85 (0.69 - 1.06)	
<i>P for trend<sup>e</sup></i>			0.12	0.09	0.14	
Former smokers	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (5.08-29.89)	164	158	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (29.90-49.90)	161	165	0.94 (0.69 - 1.28)	0.95 (0.68 - 1.31)	0.95 (0.68 - 1.32)
	3 (49.91-90.05)	178	181	0.95 (0.69 - 1.29)	0.94 (0.67 - 1.3)	0.9 (0.65 - 1.26)
	4 > 90.05	200	199	0.97 (0.71 - 1.32)	1.03 (0.73 - 1.43)	1.08 (0.77 - 1.51)
	<i>P for trend<sup>e</sup></i>			0.32	0.52	0.61
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (1.70-17.45)	110	100	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (17.46-33.35)	163	159	0.92 (0.63 - 1.32)	0.94 (0.64 - 1.38)	1.02 (0.7 - 1.51)
	3 (33.36-59.10)	196	210	0.84 (0.59 - 1.19)	0.87 (0.6 - 1.26)	0.94 (0.65 - 1.37)
	4 > 59.10	234	234	0.89 (0.62 - 1.28)	0.97 (0.66 - 1.43)	1.02 (0.7 - 1.5)
	<i>P for trend<sup>e</sup></i>			0.51	0.83	0.88
	<b>Methionine [μmol/L]<sup>d</sup></b>					
	1 (11.83-22.77)	150	132	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.78-26.16)	165	167	0.87 (0.63 - 1.2)	0.92 (0.65 - 1.3)	0.89 (0.63 - 1.27)
3 (26.17-30.47)	184	189	0.85 (0.62 - 1.16)	0.85 (0.6 - 1.18)	0.93 (0.66 - 1.32)	
4 > 30.47	204	215	0.83 (0.61 - 1.13)	0.87 (0.62 - 1.21)	0.88 (0.63 - 1.24)	
<i>P for trend<sup>e</sup></i>			0.13	0.16	0.28	
Current smokers	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (5.08-29.89)	323	297	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (29.90-49.90)	225	222	0.98 (0.75 - 1.28)	1 (0.77 - 1.3)	0.98 (0.75 - 1.28)
	3 (49.91-90.05)	150	162	0.91 (0.69 - 1.21)	0.94 (0.7 - 1.24)	0.91 (0.69 - 1.21)
	4 > 90.05	122	139	0.85 (0.63 - 1.15)	0.87 (0.65 - 1.18)	0.86 (0.64 - 1.17)
	<i>P for trend<sup>e</sup></i>			0.52	0.68	0.57
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (1.70-17.45)	340	329	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (17.46-33.35)	206	229	0.91 (0.71 - 1.17)	0.93 (0.73 - 1.2)	0.91 (0.71 - 1.17)
	3 (33.36-59.10)	157	140	1.07 (0.79 - 1.44)	1.09 (0.8 - 1.47)	1.06 (0.78 - 1.44)
4 > 59.10	117	122	1.06 (0.75 - 1.51)	1.06 (0.74 - 1.51)	1.06 (0.75 - 1.52)	

			0.97	0.95	0.98	
	<i>P for trend<sup>e</sup></i>					
	<b>Methionine [μmol/L]<sup>d</sup></b>					
	1 (11.83-22.77)	235	231	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.78-26.16)	210	200	1.01 (0.77 - 1.33)	1 (0.76 - 1.32)	1.01 (0.76 - 1.33)
	3 (26.17-30.47)	188	177	1.03 (0.76 - 1.38)	1.03 (0.76 - 1.38)	1.03 (0.76 - 1.38)
	4 > 30.47	187	212	0.85 (0.63 - 1.14)	0.84 (0.62 - 1.12)	0.84 (0.62 - 1.12)
	<i>P for trend<sup>e</sup></i>		0.39	0.32	0.36	
	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (5.08-29.89)	221	192	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (29.90-49.90)	175	174	0.89 (0.66 - 1.2)	0.91 (0.67 - 1.24)	0.87 (0.64 - 1.18)
	3 (49.91-90.05)	150	156	0.83 (0.6 - 1.13)	0.82 (0.59 - 1.14)	0.74 (0.53 - 1.03)
	4 > 90.05	129	153	0.7 (0.5 - 0.98)	0.74 (0.53 - 1.06)	0.75 (0.52 - 1.06)
	<i>P for trend<sup>e</sup></i>		0.02	0.04	0.04	
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (1.70-17.45)	185	164	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (17.46-33.35)	152	176	0.8 (0.57 - 1.11)	0.85 (0.6 - 1.18)	0.85 (0.6 - 1.19)
	3 (33.36-59.10)	172	166	0.91 (0.64 - 1.32)	0.98 (0.67 - 1.42)	0.98 (0.67 - 1.43)
	4 > 59.10	166	169	0.89 (0.6 - 1.32)	0.96 (0.64 - 1.45)	0.97 (0.64 - 1.47)
	<i>P for trend<sup>e</sup></i>		0.93	0.80	0.87	
	<b>Methionine [μmol/L]<sup>d</sup></b>					
	1 (11.83-22.77)	161	147	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.78-26.16)	161	133	1.06 (0.76 - 1.49)	1.05 (0.75 - 1.49)	1.03 (0.73 - 1.47)
	3 (26.17-30.47)	162	155	0.96 (0.69 - 1.33)	0.94 (0.67 - 1.33)	0.99 (0.69 - 1.4)
	4 > 30.47	191	240	0.69 (0.5 - 0.94)	0.69 (0.49 - 0.95)	0.68 (0.49 - 0.95)
	<i>P for trend<sup>e</sup></i>		0.02	0.01	0.03	
	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (5.08-29.89)	266	263	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (29.90-49.90)	211	213	1.02 (0.77 - 1.33)	1.03 (0.78 - 1.36)	1.03 (0.78 - 1.36)
	3 (49.91-90.05)	178	187	1 (0.76 - 1.32)	1.03 (0.78 - 1.37)	1.05 (0.79 - 1.4)
	4 > 90.05	193	185	1.12 (0.84 - 1.49)	1.15 (0.86 - 1.54)	1.16 (0.86 - 1.55)
	<i>P for trend<sup>e</sup></i>		0.52	0.34	0.35	
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (1.70-17.45)	265	265	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (17.46-33.35)	217	212	1.05 (0.8 - 1.36)	1.07 (0.82 - 1.39)	1.04 (0.8 - 1.36)
	3 (33.36-59.10)	181	184	1.01 (0.76 - 1.34)	1.03 (0.77 - 1.37)	1.01 (0.76 - 1.34)
	4 > 59.10	185	187	1.05 (0.77 - 1.42)	1.08 (0.79 - 1.48)	1.07 (0.79 - 1.46)
	<i>P for trend<sup>e</sup></i>		0.77	0.92	0.91	
	<b>Methionine [μmol/L]<sup>d</sup></b>					
	1 (11.83-22.77)	224	216	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.78-26.16)	214	234	0.86 (0.65 - 1.13)	0.88 (0.66 - 1.16)	0.89 (0.68 - 1.18)
	3 (26.17-30.47)	210	211	0.96 (0.73 - 1.28)	0.97 (0.72 - 1.3)	1 (0.75 - 1.33)
	4 > 30.47	200	187	1.06 (0.79 - 1.41)	1.07 (0.79 - 1.43)	1.07 (0.8 - 1.44)
	<i>P for trend</i>		0.85	0.86	0.81	

a) Assessed by conditional logistic regression, conditioning on individual case set and adjusting for circulating cotinine (in quartiles) and education (in 7 categories).

b) Assessed by conditional logistic regression, conditioning on individual case set and adjusting for circulating cotinine (in quartiles), education (in 7 categories), and number of years of smoking (continuous).

c) Assessed by conditional logistic regression, conditioning on individual case set and adjusting for circulating cotinine (in quartiles), education (in 7 categories), and number of pack-years of smoking (continuous).

d) Quartile cut-off points were determined based on the circulating level distribution of US matched controls.

e) P for trend assessed by the base 2 logarithm of the circulating levels.

Appendix 35- Supplementary table 2 for Article 3

Odds ratios of lung cancer for circulating levels of vitamin B6, folate and methionine among ever smoking participants in EU/AU

		European/Asian cohorts				
		Cases	Controls	Model 1 <sup>a</sup> OR (95%CI)	Model 2 <sup>b</sup> OR (95%CI)	Model 3 <sup>c</sup> OR (95%CI)
All ever smokers	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (4.37-22.41)	334	245	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.42-30.90)	216	254	0.65 (0.5 - 0.83)	0.63 (0.49 - 0.82)	0.65 (0.5 - 0.84)
	3 (30.91-44.95)	186	217	0.65 (0.5 - 0.86)	0.65 (0.49 - 0.86)	0.66 (0.5 - 0.87)
	4 > 44.96	201	221	0.71 (0.54 - 0.93)	0.71 (0.54 - 0.94)	0.69 (0.52 - 0.91)
	<i>P for trend<sup>e</sup></i>			0.31	0.65	0.17
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (0.24-7.78)	279	230	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (7.79-10.77)	220	243	0.74 (0.56 - 0.97)	0.77 (0.58 - 1.01)	0.74 (0.56 - 0.97)
	3 (10.78-15.98)	230	231	0.84 (0.64 - 1.11)	0.83 (0.63 - 1.1)	0.86 (0.65 - 1.13)
	4 > 15.98	208	233	0.76 (0.57 - 1)	0.76 (0.57 - 1.02)	0.75 (0.56 - 1)
	<i>P for trend<sup>e</sup></i>			0.06	0.06	0.04
	<b>Methionine [μmol/L]<sup>d</sup></b>					
	1 (10.63-22.81)	221	235	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.82-26.29)	249	218	1.23 (0.94 - 1.6)	1.18 (0.89 - 1.55)	1.24 (0.95 - 1.63)
3 (26.30-29.91)	235	241	0.97 (0.74 - 1.29)	1 (0.75 - 1.33)	0.99 (0.75 - 1.31)	
4 > 29.91	232	243	1.01 (0.77 - 1.33)	0.97 (0.73 - 1.3)	1.03 (0.78 - 1.37)	
<i>P for trend<sup>e</sup></i>			0.79	0.75	0.96	
Former smokers	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (4.37-22.41)	48	35	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.42-30.90)	62	55	0.8 (0.45 - 1.42)	0.88 (0.47 - 1.63)	0.83 (0.45 - 1.52)
	3 (30.91-44.95)	58	74	0.55 (0.31 - 0.98)	0.64 (0.34 - 1.2)	0.54 (0.29 - 1)
	4 > 44.96	77	81	0.66 (0.38 - 1.16)	0.67 (0.36 - 1.23)	0.61 (0.34 - 1.09)
	<i>P for trend<sup>e</sup></i>			0.72	0.51	0.28
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (0.24-7.78)	50	31	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (7.79-10.77)	53	60	0.52 (0.28 - 0.95)	0.53 (0.27 - 1.02)	0.55 (0.29 - 1.03)
	3 (10.78-15.98)	62	64	0.57 (0.32 - 1.02)	0.54 (0.29 - 1.02)	0.58 (0.31 - 1.06)
	4 > 15.98	80	90	0.51 (0.28 - 0.91)	0.48 (0.25 - 0.9)	0.5 (0.27 - 0.92)
	<i>P for trend<sup>e</sup></i>			0.03	0.02	0.01
	<b>Methionine [μmol/L]<sup>d</sup></b>					
	1 (10.63-22.81)	52	64	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.82-26.29)	56	51	1.33 (0.79 - 2.24)	1.35 (0.76 - 2.39)	1.31 (0.76 - 2.24)
3 (26.30-29.91)	68	61	1.37 (0.82 - 2.29)	1.54 (0.88 - 2.69)	1.28 (0.75 - 2.19)	
4 > 29.91	69	69	1.23 (0.75 - 2.03)	1.05 (0.61 - 1.82)	1.21 (0.72 - 2.05)	
<i>P for trend<sup>e</sup></i>			0.82	0.91	0.93	
Current smokers	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (4.37-22.41)	286	210	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.42-30.90)	154	199	0.6 (0.45 - 0.8)	0.58 (0.43 - 0.77)	0.6 (0.45 - 0.8)
	3 (30.91-44.95)	128	143	0.7 (0.51 - 0.96)	0.68 (0.49 - 0.94)	0.7 (0.51 - 0.97)
	4 > 44.96	124	140	0.74 (0.54 - 1.02)	0.76 (0.55 - 1.05)	0.74 (0.54 - 1.02)
	<i>P for trend<sup>e</sup></i>			0.33	0.44	0.31
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (0.24-7.78)	229	199	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (7.79-10.77)	167	183	0.79 (0.59 - 1.08)	0.82 (0.6 - 1.11)	0.78 (0.58 - 1.07)
	3 (10.78-15.98)	168	167	0.91 (0.67 - 1.26)	0.91 (0.66 - 1.25)	0.92 (0.67 - 1.27)
4 > 15.98	128	143	0.83 (0.59 - 1.16)	0.84 (0.6 - 1.18)	0.82 (0.59 - 1.15)	

<i>P for trend<sup>e</sup></i>			0.39	0.43	0.36
<b>Methionine [<math>\mu\text{mol/L}</math>]<sup>d</sup></b>					
1 (10.63-22.81)	169	171	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (22.82-26.29)	193	167	1.17 (0.85 - 1.6)	1.13 (0.82 - 1.55)	1.19 (0.87 - 1.64)
3 (26.30-29.91)	167	180	0.84 (0.61 - 1.17)	0.85 (0.61 - 1.19)	0.88 (0.63 - 1.23)
4 > 29.91	163	174	0.92 (0.66 - 1.28)	0.92 (0.66 - 1.29)	0.95 (0.68 - 1.33)
<i>P for trend<sup>e</sup></i>			0.60	0.66	0.81

<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
1 (4.37-22.41)	195	146	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (22.42-30.90)	140	162	0.67 (0.49 - 0.92)	0.68 (0.49 - 0.94)	0.7 (0.5 - 0.97)
3 (30.91-44.95)	123	133	0.73 (0.51 - 1.03)	0.76 (0.53 - 1.09)	0.74 (0.51 - 1.06)
4 > 44.96	117	134	0.72 (0.51 - 1.02)	0.74 (0.52 - 1.06)	0.71 (0.5 - 1.02)
<i>P for trend<sup>e</sup></i>			0.29	0.37	0.21
<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
1 (0.24-7.78)	158	121	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (7.79-10.77)	135	145	0.71 (0.5 - 1.02)	0.72 (0.5 - 1.04)	0.71 (0.49 - 1.02)
3 (10.78-15.98)	151	161	0.73 (0.51 - 1.04)	0.69 (0.48 - 1)	0.75 (0.52 - 1.08)
4 > 15.98	131	148	0.67 (0.46 - 0.97)	0.66 (0.45 - 0.98)	0.65 (0.44 - 0.96)
<i>P for trend<sup>e</sup></i>			0.03	0.02	0.01
<b>Methionine [<math>\mu\text{mol/L}</math>]<sup>d</sup></b>					
1 (10.63-22.81)	95	100	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (22.82-26.29)	144	126	1.23 (0.84 - 1.8)	1.22 (0.82 - 1.81)	1.29 (0.87 - 1.91)
3 (26.30-29.91)	164	166	0.98 (0.67 - 1.44)	1.06 (0.71 - 1.58)	1.03 (0.7 - 1.52)
4 > 29.91	172	183	0.99 (0.68 - 1.42)	1 (0.68 - 1.46)	1.03 (0.7 - 1.5)
<i>P for trend<sup>e</sup></i>			0.53	0.65	0.73

<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
1 (4.37-22.41)	139	99	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (22.42-30.90)	76	92	0.59 (0.39 - 0.9)	0.54 (0.35 - 0.83)	0.55 (0.35 - 0.84)
3 (30.91-44.95)	63	84	0.54 (0.35 - 0.85)	0.5 (0.31 - 0.79)	0.54 (0.34 - 0.85)
4 > 44.96	84	87	0.67 (0.43 - 1.04)	0.66 (0.42 - 1.03)	0.64 (0.41 - 0.99)
<i>P for trend<sup>e</sup></i>			0.69	0.67	0.46
<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
1 (0.24-7.78)	121	109	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (7.79-10.77)	85	98	0.76 (0.5 - 1.16)	0.82 (0.53 - 1.25)	0.78 (0.51 - 1.19)
3 (10.78-15.98)	79	70	1.06 (0.67 - 1.68)	1.12 (0.7 - 1.79)	1.06 (0.66 - 1.69)
4 > 15.98	77	85	0.88 (0.57 - 1.37)	0.9 (0.57 - 1.4)	0.9 (0.57 - 1.4)
<i>P for trend<sup>e</sup></i>			0.81	0.89	0.88
<b>Methionine [<math>\mu\text{mol/L}</math>]<sup>d</sup></b>					
1 (10.63-22.81)	126	135	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (22.82-26.29)	105	92	1.21 (0.82 - 1.78)	1.15 (0.77 - 1.71)	1.19 (0.8 - 1.76)
3 (26.30-29.91)	71	75	0.95 (0.62 - 1.45)	0.93 (0.6 - 1.44)	0.94 (0.61 - 1.45)
4 > 29.91	60	60	1.06 (0.67 - 1.67)	0.96 (0.61 - 1.54)	1.07 (0.67 - 1.7)
<i>P for trend<sup>e</sup></i>			0.83	0.99	0.80

a) Assessed by conditional logistic regression, conditioning on individual case set and adjusting for circulating cotinine (in quartiles) and education (in 7 categories).

b) Assessed by conditional logistic regression, conditioning on individual case set and adjusting for circulating cotinine (in quartiles), education (in 7 categories), and number of years of smoking (continuous).

c) Assessed by conditional logistic regression, conditioning on individual case set and adjusting for circulating cotinine (in quartiles), education (in 7 categories), and number of pack-years of smoking (continuous).

d) Quartile cut-off points were determined based on the circulating level distribution of EU/AU matched controls.

e) P for trend assessed by the base 2 logarithm of the circulating levels.

Appendix 36- Supplementary table 3 for Article 3

Odds ratios of lung cancer for circulating levels of vitamin B6, folate and methionine among ever smoking participants in Asia

		Asian cohorts				
		Cases	Controls	Model 1 <sup>a</sup> OR (95%CI)	Model 2 <sup>b</sup> OR (95%CI)	Model 3 <sup>c</sup> OR (95%CI)
All ever smokers	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (4.81-20.49)	490	391	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (20.50-30.82)	317	308	0.89 (0.71 - 1.13)	0.91 (0.72 - 1.15)	0.96 (0.76 - 1.22)
	3 (31.23-46.99)	205	259	0.68 (0.52 - 0.88)	0.72 (0.55 - 0.94)	0.73 (0.55 - 0.95)
	4 > 47.00	161	215	0.65 (0.49 - 0.87)	0.71 (0.54 - 0.95)	0.73 (0.54 - 0.97)
	<i>P for trend<sup>e</sup></i>			0.001	0.009	0.01
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (0.17-10.59)	418	348	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (10.60-15.00)	315	310	0.85 (0.68 - 1.07)	0.89 (0.71 - 1.13)	0.91 (0.72 - 1.15)
	3 (15.01-21.00)	222	274	0.73 (0.58 - 0.94)	0.77 (0.6 - 0.99)	0.81 (0.62 - 1.04)
	4 >21.01	218	241	0.83 (0.64 - 1.06)	0.89 (0.69 - 1.15)	0.89 (0.68 - 1.15)
	<i>P for trend<sup>e</sup></i>			0.05	0.26	0.22
	<b>Methionine [μmol/L]<sup>d</sup></b>					
	1 (12.32-23.17)	258	254	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (23.18-27.39)	282	267	1.12 (0.86 - 1.45)	1.22 (0.93 - 1.59)	1.19 (0.91 - 1.56)
3 (27.39-32.22)	288	312	0.93 (0.71 - 1.2)	0.98 (0.75 - 1.29)	0.94 (0.72 - 1.23)	
4 >32.22	345	340	1.03 (0.8 - 1.31)	1.08 (0.83 - 1.39)	1.08 (0.83 - 1.39)	
<i>P for trend<sup>e</sup></i>			0.87	0.99	0.83	
Former smokers	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (4.81-20.49)	34	22	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (20.50-30.82)	40	38	0.67 (0.32 - 1.41)	0.71 (0.33 - 1.51)	0.82 (0.36 - 1.83)
	3 (31.23-46.99)	48	52	0.6 (0.3 - 1.17)	0.6 (0.3 - 1.2)	0.63 (0.31 - 1.31)
	4 > 47.00	54	64	0.55 (0.29 - 1.06)	0.64 (0.33 - 1.26)	0.66 (0.33 - 1.32)
	<i>P for trend<sup>e</sup></i>			0.29	0.58	0.65
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (0.17-10.59)	60	37	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (10.60-15.00)	32	48	0.42 (0.23 - 0.78)	0.47 (0.25 - 0.88)	0.47 (0.25 - 0.91)
	3 (15.01-21.00)	34	45	0.45 (0.24 - 0.86)	0.53 (0.27 - 1.02)	0.61 (0.31 - 1.2)
	4 >21.01	50	46	0.65 (0.36 - 1.19)	0.82 (0.43 - 1.55)	0.85 (0.45 - 1.62)
	<i>P for trend<sup>e</sup></i>			0.26	0.74	0.76
	<b>Methionine [μmol/L]<sup>d</sup></b>					
	1 (12.32-23.17)	62	48	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (23.18-27.39)	37	41	0.66 (0.36 - 1.22)	0.74 (0.4 - 1.4)	0.77 (0.4 - 1.47)
3 (27.39-32.22)	43	45	0.69 (0.38 - 1.25)	0.74 (0.4 - 1.36)	0.71 (0.38 - 1.32)	
4 >32.22	34	42	0.59 (0.32 - 1.1)	0.61 (0.32 - 1.15)	0.56 (0.28 - 1.09)	
<i>P for trend<sup>e</sup></i>			0.17	0.17	0.17	
Current smokers	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (4.81-20.49)	456	369	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (20.50-30.82)	277	270	0.93 (0.73 - 1.18)	0.93 (0.73 - 1.19)	0.98 (0.76 - 1.26)
	3 (31.23-46.99)	157	207	0.68 (0.51 - 0.9)	0.74 (0.55 - 0.99)	0.73 (0.55 - 0.99)
	4 > 47.00	107	151	0.68 (0.49 - 0.94)	0.7 (0.5 - 0.98)	0.73 (0.52 - 1.02)
	<i>P for trend<sup>e</sup></i>			0.001	0.006	0.009
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (0.17-10.59)	358	311	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (10.60-15.00)	283	262	0.96 (0.75 - 1.23)	0.98 (0.76 - 1.26)	1 (0.78 - 1.29)
	3 (15.01-21.00)	188	229	0.8 (0.61 - 1.05)	0.82 (0.62 - 1.07)	0.84 (0.64 - 1.11)
4 >21.01	168	195	0.84 (0.64 - 1.11)	0.85 (0.64 - 1.13)	0.86 (0.65 - 1.15)	

	<i>P for trend<sup>e</sup></i>		0.09	0.18	0.18	
	<b>Methionine [<math>\mu\text{mol/L}</math>]<sup>d</sup></b>					
	1 (12.32-23.17)	196	206	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (23.18-27.39)	245	226	1.28 (0.96 - 1.72)	1.37 (1.02 - 1.86)	1.33 (0.98 - 1.8)
	3 (27.39-32.22)	245	267	1.02 (0.76 - 1.36)	1.07 (0.79 - 1.45)	1.02 (0.75 - 1.38)
	4 >32.22	311	298	1.16 (0.88 - 1.52)	1.21 (0.91 - 1.6)	1.22 (0.92 - 1.61)
	<i>P for trend<sup>e</sup></i>		0.71	0.58	0.44	
	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (4.81-20.49)	473	380	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (20.50-30.82)	293	292	0.89 (0.7 - 1.12)	0.89 (0.7 - 1.13)	0.96 (0.75 - 1.22)
	3 (31.23-46.99)	183	233	0.69 (0.53 - 0.9)	0.72 (0.55 - 0.95)	0.74 (0.56 - 0.98)
	4 > 47.00	151	195	0.69 (0.51 - 0.92)	0.75 (0.56 - 1.01)	0.77 (0.56 - 1.04)
	<i>P for trend<sup>e</sup></i>		0.003	0.02	0.04	
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (0.17-10.59)	400	328	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (10.60-15.00)	296	292	0.84 (0.66 - 1.05)	0.88 (0.69 - 1.12)	0.89 (0.7 - 1.13)
	3 (15.01-21.00)	208	250	0.75 (0.58 - 0.96)	0.8 (0.62 - 1.04)	0.82 (0.63 - 1.07)
	4 >21.01	196	230	0.76 (0.59 - 0.98)	0.82 (0.63 - 1.07)	0.81 (0.62 - 1.06)
	<i>P for trend<sup>e</sup></i>		0.02	0.15	0.09	
	<b>Methionine [<math>\mu\text{mol/L}</math>]<sup>d</sup></b>					
	1 (12.32-23.17)	225	225	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (23.18-27.39)	261	250	1.11 (0.85 - 1.46)	1.22 (0.92 - 1.61)	1.19 (0.89 - 1.58)
	3 (27.39-32.22)	279	301	0.94 (0.72 - 1.24)	1.01 (0.76 - 1.33)	0.96 (0.72 - 1.27)
	4 >32.22	335	324	1.06 (0.82 - 1.38)	1.12 (0.86 - 1.46)	1.12 (0.86 - 1.47)
	<i>P for trend<sup>e</sup></i>		0.97	0.88	0.67	
	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (4.81-20.49)	17	11	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (20.50-30.82)	24	16	0.71 (0.22 - 2.26)	0.82 (0.25 - 2.7)	0.78 (0.24 - 2.51)
	3 (31.23-46.99)	22	26	0.37 (0.12 - 1.17)	0.47 (0.14 - 1.57)	0.4 (0.12 - 1.3)
	4 > 47.00	10	20	0.29 (0.08 - 1.04)	0.32 (0.09 - 1.18)	0.31 (0.08 - 1.14)
	<i>P for trend<sup>e</sup></i>		0.04	0.07	0.05	
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (0.17-10.59)	18	20	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (10.60-15.00)	19	18	1.3 (0.39 - 4.33)	1.63 (0.46 - 5.79)	1.66 (0.46 - 5.99)
	3 (15.01-21.00)	14	24	0.65 (0.22 - 1.91)	0.59 (0.19 - 1.84)	0.78 (0.25 - 2.41)
	4 >21.01	22	11	3 (0.9 - 9.93)	3.96 (1.03 - 15.18)	3.86 (1.05 - 14.21)
	<i>P for trend<sup>e</sup></i>		0.17	0.16	0.11	
	<b>Methionine [<math>\mu\text{mol/L}</math>]<sup>d</sup></b>					
	1 (12.32-23.17)	33	29	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (23.18-27.39)	21	17	1.21 (0.44 - 3.34)	1.35 (0.48 - 3.79)	1.27 (0.45 - 3.56)
	3 (27.39-32.22)	9	11	0.64 (0.2 - 2.05)	0.7 (0.22 - 2.24)	0.75 (0.23 - 2.42)
	4 >32.22	10	16	0.51 (0.2 - 1.33)	0.54 (0.2 - 1.44)	0.54 (0.2 - 1.44)
	<i>P for trend<sup>e</sup></i>		0.34	0.42	0.38	

a) Assessed by conditional logistic regression, conditioning on individual case set and adjusting for circulating cotinine (in quartiles) and education (in 7 categories).

b) Assessed by conditional logistic regression, conditioning on individual case set and adjusting for circulating cotinine (in quartiles), education (in 7 categories), and number of years of smoking (continuous).

c) Assessed by conditional logistic regression, conditioning on individual case set and adjusting for circulating cotinine (in quartiles), education (in 7 categories), and number of pack-years of smoking (continuous).

d) Quartile cut-off points were determined based on the circulating level distribution of Asian matched controls.

e) P for trend assessed by the base 2 logarithm of the circulating levels.



Appendix 37- Supplementary table 4 for Article 3

Odds ratios of lung cancer for circulating levels of vitamin B6, folate and methionine among ever smoking participants in LC3

		All cohorts				
		Cases	Controls	Model 1 <sup>a</sup> OR (95%CI)	Model 2 <sup>b</sup> OR (95%CI)	Model 3 <sup>c</sup> OR (95%CI)
All ever smokers	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (4.37-23.93)	1300	1104	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (23.94-37.05)	911	962	0.83 (0.73 - 0.94)	0.85 (0.74 - 0.97)	0.87 (0.76 - 0.99)
	3 (37.06-62.50)	724	795	0.81 (0.7 - 0.93)	0.83 (0.72 - 0.96)	0.82 (0.71 - 0.95)
	4 > 62.51	698	772	0.78 (0.67 - 0.91)	0.83 (0.71 - 0.97)	0.81 (0.7 - 0.95)
	<i>P for trend<sup>e</sup></i>			0.003	0.03	0.02
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (0.17-10.92)	1154	1039	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (10.93-17.89)	948	957	0.91 (0.8 - 1.04)	0.93 (0.81 - 1.06)	0.94 (0.82 - 1.08)
	3 (17.90-34.92)	776	855	0.85 (0.74 - 0.98)	0.88 (0.76 - 1.02)	0.89 (0.77 - 1.03)
	4 > 34.97	755	782	0.87 (0.73 - 1.03)	0.92 (0.77 - 1.1)	0.91 (0.76 - 1.08)
	<i>P for trend<sup>e</sup></i>			0.03	0.15	0.10
	<b>Methionine [μmol/L]<sup>d</sup></b>					
	1 (10.63-22.89)	861	855	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.90-27.40)	902	849	1.07 (0.93 - 1.23)	1.09 (0.94 - 1.25)	1.09 (0.94 - 1.25)
3 (27.40-31.00)	913	913	0.99 (0.86 - 1.14)	1.01 (0.87 - 1.17)	1.02 (0.88 - 1.18)	
4 > 31.01	957	1016	0.95 (0.82 - 1.09)	0.96 (0.83 - 1.11)	0.96 (0.83 - 1.11)	
<i>P for trend<sup>e</sup></i>			0.25	0.24	0.44	
Former smokers	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (4.37-23.93)	210	163	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (23.94-37.05)	266	279	0.74 (0.57 - 0.97)	0.77 (0.58 - 1.01)	0.77 (0.58 - 1.03)
	3 (37.06-62.50)	270	288	0.72 (0.55 - 0.94)	0.76 (0.57 - 1.01)	0.7 (0.53 - 0.94)
	4 > 62.51	378	394	0.73 (0.56 - 0.95)	0.79 (0.6 - 1.04)	0.77 (0.58 - 1.02)
	<i>P for trend<sup>e</sup></i>			0.17	0.41	0.31
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (0.17-10.92)	211	164	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (10.93-17.89)	200	215	0.69 (0.52 - 0.93)	0.74 (0.54 - 1.01)	0.74 (0.54 - 1)
	3 (17.90-34.92)	275	281	0.69 (0.51 - 0.92)	0.77 (0.56 - 1.05)	0.79 (0.57 - 1.08)
	4 > 34.97	438	464	0.63 (0.47 - 0.86)	0.73 (0.53 - 1.01)	0.72 (0.52 - 1)
	<i>P for trend<sup>e</sup></i>			0.05	0.26	0.18
	<b>Methionine [μmol/L]<sup>d</sup></b>					
	1 (10.63-22.89)	263	248	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (22.90-27.40)	279	267	0.98 (0.77 - 1.26)	1.01 (0.78 - 1.31)	1.01 (0.77 - 1.31)
3 (27.40-31.00)	298	300	0.93 (0.73 - 1.18)	0.96 (0.75 - 1.24)	0.99 (0.76 - 1.28)	
4 > 31.01	284	309	0.86 (0.67 - 1.1)	0.86 (0.66 - 1.12)	0.87 (0.67 - 1.13)	
<i>P for trend<sup>e</sup></i>			0.09	0.08	0.15	
Current smokers	<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
	1 (4.37-23.93)	1090	941	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (23.94-37.05)	645	683	0.85 (0.74 - 0.99)	0.87 (0.75 - 1.01)	0.89 (0.76 - 1.03)
	3 (37.06-62.50)	454	507	0.83 (0.7 - 0.98)	0.85 (0.72 - 1.01)	0.85 (0.72 - 1.01)
	4 > 62.51	320	378	0.78 (0.64 - 0.95)	0.81 (0.67 - 0.99)	0.8 (0.65 - 0.97)
	<i>P for trend<sup>e</sup></i>			0.005	0.03	0.01
	<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
	1 (0.17-10.92)	943	875	1.00 (reference)	1.00 (reference)	1.00 (reference)
	2 (10.93-17.89)	748	742	0.96 (0.83 - 1.12)	0.97 (0.83 - 1.13)	0.98 (0.84 - 1.14)
	3 (17.90-34.92)	501	574	0.88 (0.74 - 1.04)	0.89 (0.75 - 1.05)	0.9 (0.76 - 1.06)
4 > 34.97	317	318	0.97 (0.78 - 1.21)	1.01 (0.8 - 1.26)	0.98 (0.78 - 1.23)	

<i>P for trend<sup>e</sup></i>			0.15	0.26	0.20
<b>Methionine [<math>\mu\text{mol/L}</math>]<sup>d</sup></b>					
1 (10.63-22.89)	598	607	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (22.90-27.40)	623	582	1.12 (0.94 - 1.32)	1.12 (0.95 - 1.33)	1.12 (0.95 - 1.33)
3 (27.40-31.00)	615	613	1.02 (0.86 - 1.21)	1.03 (0.86 - 1.22)	1.03 (0.87 - 1.23)
4 > 31.01	673	707	0.99 (0.84 - 1.17)	1 (0.84 - 1.19)	1 (0.84 - 1.19)
<i>P for trend<sup>e</sup></i>			0.72	0.72	0.92

<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
1 (4.37-23.93)	952	789	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (23.94-37.05)	593	655	0.76 (0.65 - 0.89)	0.79 (0.67 - 0.92)	0.82 (0.7 - 0.96)
3 (37.06-62.50)	461	497	0.81 (0.68 - 0.96)	0.84 (0.7 - 1.01)	0.82 (0.69 - 0.99)
4 > 62.51	344	409	0.68 (0.56 - 0.84)	0.72 (0.59 - 0.89)	0.7 (0.57 - 0.87)
<i>P for trend<sup>e</sup></i>			<.0001	0.001	0.001
<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
1 (0.17-10.92)	799	693	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (10.93-17.89)	677	687	0.88 (0.75 - 1.03)	0.9 (0.77 - 1.06)	0.92 (0.78 - 1.08)
3 (17.90-34.92)	489	562	0.78 (0.66 - 0.94)	0.82 (0.68 - 0.99)	0.84 (0.7 - 1.01)
4 > 34.97	385	408	0.78 (0.61 - 0.98)	0.86 (0.67 - 1.1)	0.82 (0.64 - 1.05)
<i>P for trend<sup>e</sup></i>			0.005	0.05	0.02
<b>Methionine [<math>\mu\text{mol/L}</math>]<sup>d</sup></b>					
1 (10.63-22.89)	476	464	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (22.90-27.40)	538	490	1.11 (0.92 - 1.33)	1.14 (0.94 - 1.38)	1.13 (0.93 - 1.37)
3 (27.40-31.00)	618	624	0.97 (0.81 - 1.17)	1.02 (0.84 - 1.23)	1 (0.83 - 1.21)
4 > 31.01	718	772	0.92 (0.77 - 1.1)	0.94 (0.79 - 1.13)	0.94 (0.78 - 1.13)
<i>P for trend<sup>e</sup></i>			0.16	0.17	0.32

<b>Vitamin B6 (Pyridoxal 5'-phosphate) [nmol/L]<sup>d</sup></b>					
1 (4.37-23.93)	348	315	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (23.94-37.05)	318	307	0.96 (0.77 - 1.2)	0.98 (0.78 - 1.23)	0.96 (0.76 - 1.2)
3 (37.06-62.50)	263	298	0.81 (0.64 - 1.03)	0.83 (0.65 - 1.06)	0.82 (0.65 - 1.05)
4 > 62.51	354	363	0.94 (0.74 - 1.19)	0.98 (0.77 - 1.25)	0.97 (0.77 - 1.24)
<i>P for trend<sup>e</sup></i>			0.96	0.71	0.91
<b>Folate (Vitamin B9) [nmol/L]<sup>d</sup></b>					
1 (0.17-10.92)	355	346	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (10.93-17.89)	271	270	0.98 (0.77 - 1.25)	0.98 (0.76 - 1.25)	0.97 (0.76 - 1.24)
3 (17.90-34.92)	287	293	0.99 (0.78 - 1.28)	1 (0.77 - 1.28)	1 (0.77 - 1.28)
4 > 34.97	370	374	1.01 (0.78 - 1.31)	1.03 (0.79 - 1.34)	1.03 (0.79 - 1.33)
<i>P for trend<sup>e</sup></i>			0.89	0.93	0.90
<b>Methionine [<math>\mu\text{mol/L}</math>]<sup>d</sup></b>					
1 (10.63-22.89)	385	391	1.00 (reference)	1.00 (reference)	1.00 (reference)
2 (22.90-27.40)	364	359	1.02 (0.83 - 1.26)	1.02 (0.82 - 1.26)	1.03 (0.83 - 1.28)
3 (27.40-31.00)	295	289	1.03 (0.82 - 1.29)	1.01 (0.8 - 1.27)	1.05 (0.83 - 1.33)
4 > 31.01	239	244	1.02 (0.8 - 1.29)	1.02 (0.8 - 1.31)	1.02 (0.8 - 1.3)
<i>P for trend<sup>e</sup></i>			0.97	0.97	0.93

a) Assessed by conditional logistic regression, conditioning on individual case set and adjusting for circulating cotinine (in quartiles) and education (in 7 categories).

b) Assessed by conditional logistic regression, conditioning on individual case set and adjusting for circulating cotinine (in quartiles), education (in 7 categories), and number of years of smoking (continuous).

c) Assessed by conditional logistic regression, conditioning on individual case set and adjusting for circulating cotinine (in quartiles), education (in 7 categories), and number of pack-years of smoking (continuous).

d) Quartile cut-off points were determined based on the circulating level distribution of LC3 matched controls.

e) P for trend assessed by the base 2 logarithm of the circulating levels.

