Maastricht University

Big data and prevention of bladder cancer

Citation for published version (APA):

Yu, Y. (2020). *Big data and prevention of bladder cancer*. [Doctoral Thesis, Maastricht University]. ProefschriftMaken Maastricht. https://doi.org/10.26481/dis.20200707ey

Document status and date:
Published: 01/01/2020

DOI:
10.26481/dis.20200707ey

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 23 Apr. 2024

Big Data and Prevention of Bladder Cancer

Evan Yi-Wen Yu

The work presented in this thesis was conducted at the Section of Complex Genetics and Epidemiology, Department of Genetics and Cell Biology, School for Nutrition and Translational Research in Metabolism (NUTRIM), Faculty of Health, Medicine and Life Sciences at Maastricht University.

ISBN: 978-94-6380-820-0

Cover design: Evan Y.W. Yu, Shuxin Liu
Layout and printed by: ProefschriftMaken // www.proefschriftmaken.nl

Copyright © 2020 E. Yu, Maastricht, The Netherlands all rights reserved. No part of this thesis may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage or retrieval system, without permission in writing from the author, or, when appropriate, from the publishers of the publications.

Big Data and Prevention of Bladder Cancer

THESIS

to obtain the degree of Doctor at Maastricht University,

on the authority of the Rector Magnificus,

Prof. dr. Rianne M. Letschert

in accordance with the decision of the Board of Deans,

to be defended in public on

Tuesday, July 7th, 2020

at 16:00 hours

by

Evan Yi-Wen Yu

Born in Dongshan P.R. China, 壹玖玖零

Promotiecommissie

Promotorte: Prof. dr. M.P. Zeegers

Co-promotoren: Dr. A. Wesselius

Dr. S. Mehrkanoon

Beoordelingscommissie:

Prof. dr. L. J.M. Smits (Chiarman)

Prof. dr. C. La Vecchia (University of Milan)

Prof. dr. F. J. van Schooten

Dr. T. Marcelissen

Dr. I. Kanera (Zuyd University of Applied Sciences)

献给我的父亲

TABLE OF CONTENTS

List of Figures 9

List of Tables 11

Abbreviations 15

Chapter 1 General Introduction 19

Chapter 2 Health-Related Quality of Life in Bladder Cancer Patients around the Time of Diagnosis 45

Chapter 3 The Association between Coffee Consumption and Bladder Cancer Risk: A Pooled Analysis from 13 Case-control Studies 69

Supplementary Materials to Chapter 3 99

Chapter 4 Coffee Consumption and Bladder Cancer Risk: A Pooled Analysis of 501,604 Participants from 12 Prospective Cohort Studies 105

Supplementary Materials to Chapter 4 142

Chapter 5 Vegetable Intake and Bladder Cancer Risk: A Pooled Analysis of 580,953 Participants from 14 Prospective Cohort Studies 159

Supplementary Materials to Chapter 5 201

Chapter 6 Grain and Dietary Fibre Intake and Bladder Cancer Risk: A Pooled Analysis of 574,726 Participants from 13 Prospective Cohort Studies 221

Supplementary Materials to Chapter 6 271

Chapter 7 A Machine Learning Approach to Investigate Food Groups related to Bladder Cancer Risk based on 18 Case-control and 5 Prospective Cohort Studies 287

Chapter 8 Summary and General Discussion 317

Additional Information 345

Samenvatting 349

Valorisation Addendum 353

List of Publications & Presentations 357

About the Author 361

Acknowledgements 363

LIST OF FIGURES

Figure 1-1 Anatomical Features of the Urinary Bladder a) Sider View and b) Front View 22

Figure 1-2 Illustration of the Bladder Composition and Tumour Invasion by Stage (adapted with permission from Springer Nature) 24

Figure 1-3 Map of the Study BCPP Sites across the West-Midlands 32

Figure 1-4 Map of the Consortium BLEND Study Centres across the World 34

Figure 3-1 Forest Plot of Meta-Analysis with ORs and 95% CIs for Ever and Never Coffee Consumption with Bladder Cancer Risk Adjusted for Age, Sex and Smoking by Geographical Regions 82

Figure 3-2 Dose-response Relationships between Coffee Consumption and the Risk of Bladder Cancer among a) Overall study population; b) Ever Smokers; c) Never Smokers 83

Figure 4-1 Dose-response Relationships between Coffee Consumption and the Risk of Bladder Cancer among a) Overall Study Population; b) Ever Smokers; c) Never Smokers; d) Males; e) Male Smokers 127

Figure 6-1 Dose-response Relationships between Grains/Dietary Fibre Intake and the Risk of Bladder Cancer among a) Total Whole Grain; b) Total Refined Grain; c) Total Dietary Fibre; d) Cereal Fibre 255

Figure 7-1 Decision Tree Derived after Applying C5.0 295

Figure 7-2 Importance Values of Input Variables after C5.0 297

LIST OF TABLES

Table 1-1 Tumour, Node, Metastasis (TNM) Classification System for Bladder Cancer 23

Table 1-2 Summary of Association Between Foods, Nutrients and Dietary Patterns and Bladder Cancer Risk 27

Table 2-1 Socio-demographic Characteristics of Study Participants 52

Table 2-2 HRQoL as Measured by the EORTC QLQ-C30 for NMIBC and MIBC Separately 53

Table 2-3 Multivariate Linear Regression of Age, Sex, Stage and Comorbidity on Each HRQoL Domain 55

Table 2-4 Physical and Mental Health Scores across the Different Socio-Demographic Variables 57

Table 3-1 Characteristics of the Study Population (5,911 Cases and 16,172 Controls) and Coffee Consumption Category According to Age, Sex and Smoking Status 77

Table 3-2 Adjusted Odds Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Level Stratified by Smoking 79

Table 4-1 Characteristics of the Study Population (2,601 cases and 499,003 non-cases) and Coffee Consumption Amount of Included Studies 112

Table 4-2 Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Amount Stratified by NMIBC and MIBC 117

Table 4-3 Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Amount Stratified by NMIBC and MIBC 119

Table 4-4 Hazard Ratios and 95% Confidence Intervals of Caffeinated and Decaffeinated Coffee Consumption with Bladder Cancer Risk Based on Model 3 [b, c, d] Stratified by NMIBC and MIBC 126

Table 5-1 Characteristics of the Study Population (3,253 Cases and 577,700 Non-cases) and Vegetable Intake of Included Studies 167

Table 5-2 Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Intake of Vegetable Stratified by Subtype, Sex and Smoking Status 174

Table 5-3 Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Intake of Non-starchy or Starchy Vegetable 179

Table 5-4 Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Intake of Green Leafy Vegetable or Cruciferous Vegetable 183

Table 5-5 Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Diet Diversity Scores of Vegetable 186

Table 6-1 Characteristics of the Study Population (3,214 Cases and 571,512 Non-cases) of Grains and Fibre Intake of Included Studies 234

Table 6-2 Risk of Bladder Cancer According to Intake of Total Grain, Total Whole Grain and Total Refined Grain 240

Table 6-3 Risk of Bladder Cancer According to Individual Intake of Whole Grains and Refined Grains 246

Table 6-4 Risk of Bladder Cancer According to Intake of Total Dietary Fibre and Individual Source of Dietary Fibre 250

Table 6-5 Joint Association of Intake of Total Whole Grain and Total Dietary Fibre with Bladder Cancer 253

Table 7-1 The Baseline Characteristics and Food Group Information of Included Studies 292

Table 7-2 Classification Rules Derived from C5.0 “Ruleset” 298

Table 8-1 Overview of the Main Methods and Findings of Each Chapter 310

Table 8-2 Mean and Standard Deviations for Intake from Long-form FFQs and Short-form FFQs with Spearman’s Correlation Coefficient 312

ABBREVIATIONS

IARC, *International Agency for Research on Cancer*

WHO, *World Health Organization*

AICR, *American Institute for Cancer Research*

CUP, *Continuous Update Project*

CIS, *carcinoma in situ*

TURBT, *transurethral resection of the bladder tumour*

TNM, *tumour, node, and metastasis*

NMIBC, *non-muscle*-invasive bladder cancer

MIBC, *muscle-invasive bladder cancer*

HRQoL, *health-related quality of life*

RCT, *randomized clinical trials*

PCA, *principle component analysis*

BCPP, *Bladder Cancer Prognoses Programme*

CRFs, *Case Report Forms*

EORTC, *European Organisation for the Research and Treatment of Cancer*

BLEND, *BLadder cancer Epidemiology and Nutritional Determinants*

EPIC, *European Prospective Investigation into Cancer and Nutrition*

CFA, *confirmatory factor analysis*

CRUK, *Cancer Research UK Cancer*

SD, *standard deviation*

FFQ, *food frequency questionnaire*

RMSEA, *root mean squared error of approximation*

OR, *odds ratio*

CI, *confidence interval*

POR, *pooled odds ratio*

CYP1A2, *Cytochrome P450 1A2*

PAHs, *polycyclic aromatic hydrocarbons*

ATM, *ataxia-telangiectasia mutated*

PAR, *population-attributable risk*

BMI, *body mass index*

SES, *socioeconomic status*

GLS, *generalized least squares*

NLCS, *NetherLands Cohort Study*

VITAL, *VITamins and lifestyle cohort study*

RERF-LSS, *Radiation Effects Research Foundation-Life Span Study*

ICD-O, *International Classiication of Diseases for Oncology*

HR, *hazard ratio*

ml, *millilitre*

kcal, *kilocalorie*

g, *gram*

mg, *microgram*

MCCS, *Melbourne Collaborative Cohort Study*

DDS, *diet diversity score*

DNA, *deoxyribonucleic acid*

USDA, *United States Department of Agriculture*

MAR, *missing at random*

MCAR, *missing completely at random*

MNAR, *missing not at random*

TPR, *true positive rate*

FRP, *false positive rate* (FPR)

ROC*: receiver operating curve*

AUC*: area under the curve*

RCS, *restricted cubic spline*

AR, *alkyl resorcinol*

***SCC**, Spearman's correlation coefficient*

Chapter 1

General Introduction

1.1 Nutrition

1.1a Nutritional epidemiology

Since ancient times, observations of the relations between food choices and health have been made, but nutritional epidemiology, which can be regarded as the science of systematically studying these relations in human populations, is one of the pivotal disciplines in epidemiology and has played a key role in official nutritional guidance in recent years. Whereas in the past the goal of nutritional guidance was only on the prevention of typical deficiency syndromes (*e.g.* protein energy malnutrition, iron deficiency anaemia), it now also aims to prevent and control morbidity and premature mortality resulting from chronic diseases. Due to this evolving purpose of nutritional guidance, the field of nutritional epidemiology has emerged exponentially during the last decade. Although epidemiology is particularly suitable to study the relation between diet/nutrition and chronic diseases, the complex nature of diet has, until yet, limited the capability of nutritional epidemiology to find strong evidence for causal associations. Diet not only represents an unusually complex set of exposures that are strongly intercorrelated, which makes it extremely difficult to examine, it is also often difficult to investigate because of the limited range of variation within a study population, especially when examined within one country. In addition, the associations between dietary intake and cancer is often disentangled by the impact of diet on other modifiable lifestyle factors such as physical activity.

1.1b Nutritional epidemiology, cancer and public health

The *International Agency for Research on Cancer* (IARC) estimated that since 2010 more people are dying of cancer than of cardiovascular diseases, making cancer the number one cause of death worldwide [1]; this high mortality of cancer will double by 2020 and triple by 2030 [2]. The prevention of cancer, therefore, presents a major challenge and solution to a growing global public health problem. It is likely that diet accounts for much of the cancer morbidity and mortality although it is not yet clear which dietary factors cause which tumours. In 1997 the *World Cancer Research Fund* (WCRF)/*American Institute for Cancer Research* (AICR) initiated a *Continuous Update Project* (CUP) by making use of up-to-date source of scientific research on cancer prevention and survivorship through diet, nutrition, physical activity and cancer [3], and in 2004 the *World Health Organization* (WHO) developed a Global Strategy on Diet, Physical Activity and Health [4]. One of their main objectives was to increase

the understanding of the influences of diet and cancer, as more knowledge on carcinogenesis is needed to continually improve prevention strategies.

1.1c Nutrition, bladder cancer and public health

Bladder cancer is among the top ten most common cancer types in the world, with approximately 550,000 new cases (3.0% of all new cancer diagnoses) and 200,000 deaths (2.1% of all cancer deaths) annually [5, 6]. A strong male predominance of bladder cancer is estimated with respective incidence and mortality rates of 9.6 and 3.2 per 100,000 in males (about 4 times of those of females), making this type of cancer fourth most common cancer and eighth leading cause of cancer death in males [5-8]. As with many solid tumours, incidence of bladder cancer increases with age and rarely occurs before the age of 40-50 years (the average age of people diagnosed with bladder cancer is 73 years old) [9, 10]. The incidence of bladder cancer varies significantly between geographical regions, with the highest observed in Europe and North America [11]. As suggested for many tumours, the development of bladder cancer is also likely to be influenced by diet. The role of diet in bladder cancer could be particularly pronounced because the bladder is an excretory organ. Given that the bladder is constantly exposed to both harmful and favourable components of a diet that are excreted through the urinary tract and come in direct contact with the bladder lining [12], it is, therefore, not surprising that the role of dietary intake in bladder cancer prevention is now widely recognized [13].

1.2 The Bladder

1.2a The anatomy and biological function of the bladder

The urinary bladder is a muscular sac in the pelvis, just below the kidneys and right behind the pubic bone. It is divided into four anatomical parts; the **apex** (located superiorly, pointing towards the pubic symphysis, connected to the umbilicus by the median umbilical ligament), **body** (main part of the bladder, located between the apex and the fundus), **fundus** (located posteriorly, triangular-shaped with the tip of the triangle pointing backwards), and **neck** (formed by the convergence of the fundus and the two inferolateral surfaces, and continuous with the urethra) [14, 15]. The main functions of bladder are a) temporary storage of urine; the bladder is a hollow muscular organ with distensible walls, allowing it to accommodate up to 300-400ml of urine in healthy adults; b) assists in the expulsion of urine; the musculature

of the bladder contracts during micturition (*i.e.* action of urinating), with concomitant relaxation of the sphincters. Urine enters the bladder through the left and right ureters and exits via the urethra. As such the bladder wall is consistently exposed and contacted to urine, in which any compounds would make effects on bladder throughout the micturition cycle [16] (**Figure 1-1**).

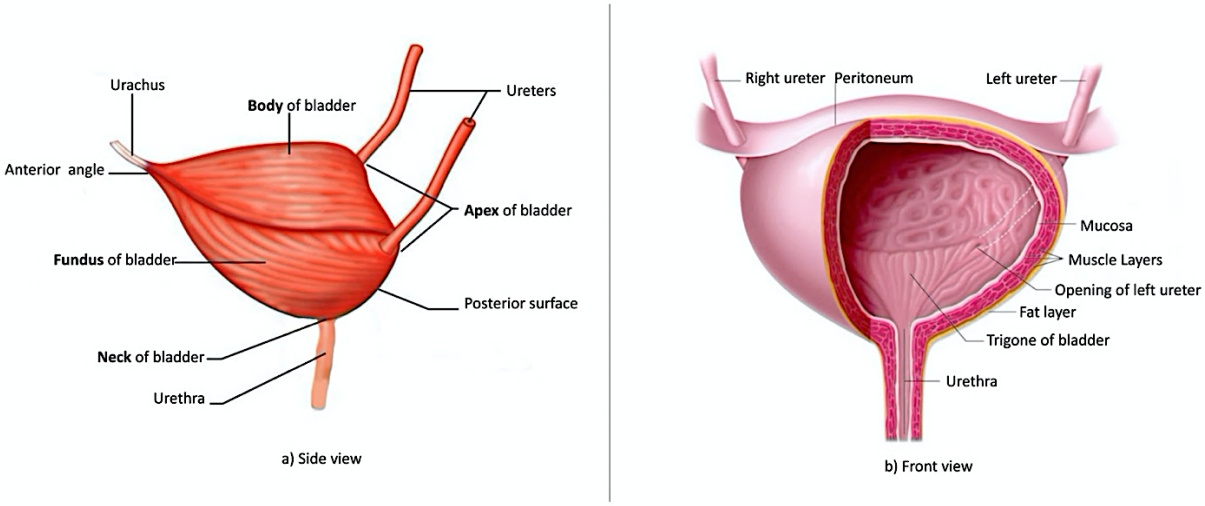


Figure 1-1 *Anatomical Features of the Urinary Bladder a) Sider View and b) Front View [Adapted from a) "Structure and development of urinary bladder" by Dr. Tatheer: https://dokumen.tips/documents/development-structure-of-urinary-bladder-dr-tatheer-zahra-assistant-professor.html; b) European Association Urology (EAU) guideline: https://uroweb.org/american-society-of-clinical-oncology-endorses-eau-guideline-on-muscle-invasive-and-metastatic-bladder-cancer/]*

1.3 Bladder Cancer

1.3a Histopathology, diagnosis and tumour stages of bladder cancer

Bladder cancer presents mainly as a papillary or non-papillary growth into the bladder lumen, a flat lesion, or as a solid tumour growing into the bladder wall. The vast majority (over 90%) of bladder cancer originates from the urothelial cell layer (*i.e.* transitional cell) and is thus defined as urothelial carcinoma. The remaining 10% of all bladder cancers have other cellular origins, *e.g.* the glandular cells (adenocarcinoma) or the bladder lining (squamous cell carcinoma) [17].

The most common symptom of bladder cancer is visible or non-visible haematuria (*i.e.* blood in the urine). Other symptoms of bladder cancer diagnosis include dysuria, bladder pain, frequent urination or sudden urge, and frequent urinary tract infections [18]. In order to diagnose bladder cancer, patients experiencing one or more of these symptoms are subjected to a) cystoscopy, in which the interior of the bladder is inspected; and b) voided urinary cytology, in which cells collected from a urine are examined [19, 20]. Subsequently, collection of tissue is performed during a *transurethral resection of the bladder tumour* (TURBT) [21]. The pathological evaluation is performed on the resected TURBT sample to determine the stage of the tumour. Staging of a tumour describes where the tumour is located, how deep the tumour has grown into the bladder wall, and whether it has extended to other organs. For bladder cancer, staging employs the *tumour, node, and metastasis* (TNM) system, where T describes the primary tumour extent, N describes the presence and extent of regional nodal metastases, and M reflects distant metastatic cancer [22] (**Table 1-1**).

Table 1-1 *Tumour, Node, Metastasis (TNM) Classification System for Bladder Cancer*

T-Primary Tumour	
T Categories	**T Criteria**
Ta	Non-invasive papillary carcinoma
Tis	*Carcinoma in situ* (CIS): “flat tumour”
T1	Tumour invades subepithelial connective tissue
T2	Tumour invades muscle
T2a	Tumour invades superficial muscle (inner half)
T2b	Tumour invades deep muscle (outer half)
T3	Tumour invades peri vesical tissue:
T3a	microscopically
T3b	macroscopically (extravesical mass)
T4	Tumour invades any of the following: prostate stroma, seminal vesicles, uterus, vagina, pelvic wall, abdominal wall
T4a	Tumour invades prostate stroma, seminal vesicles, uterus or vagina
T4b	Tumour invades pelvic wall or abdominal wall
TX	Primary tumour cannot be assessed

Table 1-1_continued *Tumour, Node, Metastasis (TNM) Classification System for Bladder Cancer*

T0	No evidence of primary tumour
N-Regional Lymph Nodes	
N Categories	**N Criteria**
N0	No regional lymph node metastasis
N1	Metastasis in a single lymph node in the true pelvis (hypogastric, obturator, external iliac, or presacral)
N2	Metastasis in multiple regional lymph nodes in the true pelvis (hypogastric, obturator, external iliac, or presacral)
N3	Metastasis in a common iliac lymph node(s)
NX	Regional lymph nodes cannot be assessed
M-Distant Metastasis	
M Categories	**M Criteria**
M0	No distant metastasis
M1a	Non regional lymph nodes
M1b	Other distant metastases
MX	Metastasis cannot be measured.

Bladder cancer can broadly be categorised into broad groups of *non-muscle-invasive bladder cancer* (NMIBC) *and muscle-invasive bladder cancer* (MIBC), as the level of penetration into the muscle provides most comprehensive prognosis of the outcome. NMIBC and MIBC share the topography, however, there is substantial evidence for them to be considered separate types of cancers. Current research shows they might develop in distinct pathways, papillary- and non-papillary that overlap to a degree, but clinically result in two different forms of bladder cancer [23, 24]. Approximately 70%-80% of newly diagnosed bladder cancer are NMIBCs and include tumours of stage Ta (50-70%), stage T1 (20-40%), and stage Tis/CIS 5-10% [25-27] (**Figure 1-2**).

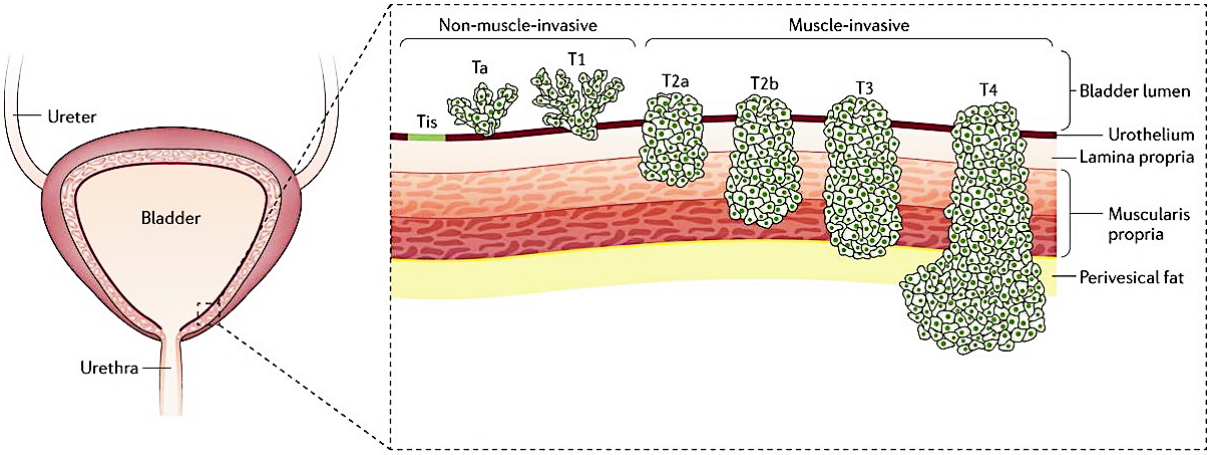


Figure 1-2 *Illustration of the Bladder Composition and Tumour Invasion by Stage (Adapted from Nature Reviews Disease Primers 2017 [28])*

1.3b Prognosis of bladder cancer

The general 5-year survival rate for people with bladder cancer is 77%, the overall 10-year survival rate is 70% and the overall 15-year survival rate is 65% [29]. The rates of survival depend on many factors, including sex, age, recurrence status, number of tumours, T stage, presence of concomitant CIS, and tumour grade. Although the prognosis for NMIBC patients is generally good, they have a high propensity for tumour recurrence (>50%) and a risk for progression to MIBC (5-20%) [26, 27, 30, 31]. The prognosis of MIBC is much worse, as it rapidly metastasizes [32]. Since the majority of bladder cancer cases are superficial tumours with a favourable prognosis (following treatment), patients need to be closely followed-up over the course of their lives. Not only does cause a significant burden of disease on an individual patient, it also adds to the overall cost of bladder cancer care resulting this type of cancer to be one of the most expensive cancers to treat on a per patient basis. Worldwide, costs per patient annually are estimated between € 80,000 to € 160,000.

1.3c Health-Related Quality of Life of bladder cancer patients

Because of the high survival rate of bladder cancer, often long-term follow-up, including invasive procedure that causes pain and discomfort, is required. Therefore, bladder cancer can be considered as a chronic disease, in which *health-related quality of life* (HRQoL) plays an important role.

HRQoL is a multidimensional concept that reflects the extent to which the psychological, physical and social wellbeing of an individual is affected by the disease and its therapy [33]. Age, sex, tumour localization, tumour size, treatment modality, symptoms on site, comorbidity, emotional status, lifestyle, marital status, income and performance status have all been shown to affect the overall HRQoL in patients with cancer. The HRQoL is used in medical research in order to understand the diseases on several aspects, including; a) quantify the impact of disease burden; b) compare patient-to-general or interpatient populations; c) assess the effectiveness of medical interventions; and d) improve patient-clinician communication. It is, therefore, of utmost importance that a clinician is aware and keeps tracking of a patient's HRQoL, in order to tailor healthcare according to the individual patient's needs and to deliver high-quality healthcare. During the last decade several researchers focussed on the HRQoL of bladder cancer patients, however, most of the studies are cross-sectional or retrospective, thereby, unable to assess the HRQoL before or around time at diagnosis, which is important for the evaluation of any changes in a patient's HRQoL due to disease or treatment [34-40].

1.3d Foods, nutrients and bladder cancer

Although smoking and exposure to certain chemicals in the working and general environments are the major risk factors for bladder cancer [41-45], evidence of an essential role for diet on the bladder cancer development is emerging. However, unless this steep increase, the literature on the influence on diet and bladder cancer risk is still considered sparse and remains inconclusive. The 2018 third expert report of the WCRF/AICR [46] , therefore, concluded that to date there is limited evidence on the association between specific foods, nutrients and the risk of bladder cancer.

So far, there is strong evidence that the inorganic contaminated foods/drinks (*e.g.* rice and water) may increase bladder cancer risk [47] and limited evidence that greater consumption of tea and of fruits and vegetables may decrease the risk of bladder cancer. Among different kind of vegetables, especially non-starchy vegetable, showed the protective effect. In addition, there is currently no evidence of any nutrient present in food or supplement to be associated with bladder cancer risk [12]. A likely reason for this mainly lacking evidence on the influence

of foods and nutrients on bladder cancer risk is that most previous studies had insufficient sample sizes and thus lacked adequate statistical power for detailed analyses.

1.3e Dietary patterns and bladder cancer

During the last decade, an increasing number of studies are taking a more holistic dietary approaches by assessing dietary patterns rather than single food/nutrient items in relation to bladder cancer risk. So far, both the dietary patterns labelled as "sweet beverages" (high loadings of coffee, tea, and added sugar) and "Western" (generally low in fresh fruits and vegetables, but high in saturated fats, red and processed meats, fried eggs, potatoes, and red wine) showed to be associated with an increased bladder cancer risk, while the "Mediterranean diet" (generally high in fruits, vegetables, legumes and cereals, moderate-to-high in fish, moderate in alcohol (mostly wine), low-to-moderate in milk and dairy products, and low in meat and meat products) showed to be protective [48-51]. No other dietary patterns have yet been linked to the development of bladder cancer (**Table 1-2**).

Table 1-2 *Summary of Association between foods, nutrients and dietary patterns and bladder cancer risk*

Dietary Intake and Bladder Cancer			
		Decreases Risk	Increases Risk
Strong Evidence	Convincing		Arsenic in drinking water
	Probable		
Limited Evidence	Limited-Suggestive	Vegetable and fruit; Tea	
	Limited-No conclusion	Cereals (grains) and their products, pulses (legumes), meat, poultry, fish, total fat, milk, yoghurt, cheese, dietetic foods, soft drinks, diet drinks, fruit juices, coffee, green tea, caffeine, alcohol, chlorinated surface water, total fluid intake, sweeteners, frying, carbohydrate, protein, vitamin A, vitamin C, serum 25-hydroxy vitamin D, vitamin E, calcium, folate, selenium, beta-carotene, alpha-carotene, lycopene, beta-cryptoxanthin, lutein, zeaxanthin, flavonoids, tocopherols, multivitamin supplements	
Substantial Effect on Risk Unlikely		No identified	
Evidence of Dietary Patterns		Limited-No conclusion	

1.4 Dietary Assessment

1.4a Epidemiological approaches for dietary assessment

Correlation or ecologic studies were the first epidemiological studies investigating the relation between diet and diseases occurrence. Within these studies the diseases rates are compared between populations with the populations per capita intake/consumption of specific dietary factors. Although these types of studies are very useful in the generation of hypotheses about dietary factors that may be associated with disease occurrence, they are considered as providing the weakest evidence for causality. In contrast, *randomized clinical trials* (RCTs) are often considered the ultimate approach to provide evidence for causality in nutritional epidemiology. However, RCTs have some methodological and practical limitations in determining effect of dietary intake of disease occurring late in life, such as cancer. For example, an RCT can only test a specific intervention, in a specific population, over a relatively short period of time. Therefore, results achieved from RCTs may not yield for a different dose, population or over a longer period of time. Moreover, intervening at a late period in life, which is necessary to ensure an adequate number of cancer cases, might not be effective, since cumulative exposures over time are likely to be important and some cancer risk factors are known to have the greatest influence early in life. To overcome these weaknesses of both correlation/ecological studies and RCTs, epidemiologists conducted a large number of case-control and cohort studies to answer nutrition question. Although their great potential in nutritional epidemiology, two major problems remain using either a case-control or cohort designs to identify causality; a) In both single case-control and cohort studies only a limited range of variation in diet can be assessed and b) most cohort studies are not be able to accumulate a sufficient number of cases within a reasonable amount of time. Therefore, large scale, especially international and prospective, studies or consortia are needed, which will transform cancer- and nutritional epidemiology into “big data” era.

1.4b Statistical approaches for dietary assessment

Broadly, there are two types of approaches which can be taken in order to assess the dietary intake on disease risk, namely; a) a reductionist approach and b) a holistic approach. Both these approaches are indispensable and of common use in nutritional epidemiology. A reductionist approach is a more classic approach and focuses on the influence of single foods or food groups on disease risk. The interpretation of the results from such an approach are

often straight forward and can be easily translated into practical health message. Although this type of analysis has been quite valuable, it also has some conceptual and methodological limitations; a) people do not consume isolated foods/nutrients; b) the variety of food items consumed could also influence disease risk; c) the high correlation among some foods and nutrients makes it difficult to examine their separate effects; d) the effect of a single nutrient may be too small to detect but the cumulative effects of multiple nutrients included in a dietary pattern may be sufficiently large to be detectable; e) analyses based on a large number of individual nutrients or food items may produce significant associations simply by chance. A more holistic approach, in which dietary patterns are explored, might therefore be more promising in terms of disease prevention. Several conventional analysis techniques are available for extracting dietary patterns including factor and cluster analyses, investigator-driven methods, such as dietary indices and dietary scores, and data-driven methods, such as *principle component analysis* (PCA). Although these techniques are widely used and might reveal some important information on the relation between dietary patterns and common cancers, they all draw subjective conclusions since they are based on series of priori assumptions, which may differ among researchers. In addition, since nutritional epidemiological studies is likely to enter the “big data” era, with its complexities in term of missing data, transformation misunderstanding, different study-design drawbacks and imperfect linear- and non-linear data, these conventional techniques lack the ability to capture, manage, or process this type of “big data”. New techniques in nutritional epidemiology are, therefore, warranted. A relatively new technique with the ability to overcome these limitations is machine learning. Although relatively new, machine learning already proved its great potential to help researchers to focus on the most important information of data without human intervention or assistance. In nutritional epidemiology this would allow epidemiologists to define unexpected dietary information that might not be recognized using conventional statistical approaches [52]. However, evidence on the appropriateness of this technique to unravel dietary patterns related to disease risk is mainly lacking. In addition, it should be acknowledged that translating the results from machine learning into simple health message will be a difficult challenge.

1.5 Outline of the Thesis

1.5a Thesis content

This thesis aims to gain insight of the HRQoL of bladder cancer patients at the time of diagnosis and to increase our understanding of the influence of dietary intake on bladder cancer risk, by making use of "big data", compromising data of over 13,000 bladder cancer cases out of over 720,000 participants and over 1,000 variables. Our first aim was tested in the West Midlands *Bladder Cancer Prognoses Programme* (BCPP) and described in **Chapter 2**. Our second aim is described in **Chapter 3-7** and can be divided in two parts. In the first part we used a hypothesis-driven approach to establish the relation between dietary intake and bladder cancer (**Chapter 3-6**), while in the second part (**Chapter 7**) an exploratory approach was used.

Coffee is one of the most world-popular beverages with potential carcinogenicity; however, evidence on the influence of coffee on bladder cancer risk is controversial. Therefore, in **Chapter 3** and **Chapter 4** we investigated the effect of coffee consumption on bladder cancer risk, by pooling data from case-control and cohort studies, respectively. Vegetables are thought to have positive health effects in terms of cancer prevention. Several epidemiological studies indeed show a preventative effect of vegetable on bladder cancer risk. However, also controversial results are shown; in addition, analyses for different subgroups and/or individual types of vegetable are mainly lacking. Therefore, in **Chapter 5**, we established the effects of vegetable intake, both total, different subgroups and types, on the risk of bladder cancer by pooled cohort studies. **Chapter 6** assessed the influences of whole- and refined grains on bladder cancer risk. In addition, also the effect of dietary fibre, as an important nutrient of grain, was assessed. Given that the examination of food combinations aims to deal with the complexity and unpredictability of dietary intake, an exploratory supervised machine learning approach was applied in **Chapter 7** to extract and prioritize the food groups related to bladder cancer. **Chapter 8** summarises all research done for this thesis by discussing; a) the results in general, b) the interpretation and implication of main findings, c) the methodological consideration, d) the statistical validity and limitations, e) the future aspects of cancer- and nutritional research. To finish, **Chapter 8** provides final conclusions for the reader of this thesis.

1.5b Studies included in this thesis

The studies presented in this thesis were embedded in two population-based studies: a prospective cohort study of bladder cancer patients (BCPP) and an international consortium of observational studies investigating bladder cancer risk (BLEND).

1.5b-1 Bladder Cancer Prognosis Programme (BCPP) study

The West Midlands *Bladder Cancer Prognosis Programme* (BCPP) is one of the largest prospective cohorts of bladder cancer patients worldwide including approximately 1,300 NMIBC and MIBC patients at baseline. The study was initiated by the Cancer Research UK Bladder Cancer Group at the University of Birmingham. The Nottingham Multi-Centre Research Ethics Committee gave approval for the study (reference: 06/MRE04/65 and clinicaltrials.gov (registration number: NCT00553215). The aims of this multi-centre study are to; a) improve knowledge on prognosis by identifying (modifiable) factors that influence recurrence or progression of bladder cancer, b) to develop a prognostic tool that could better predict these outcomes in patients and c) to better understand bladder cancer patient behaviour and self-perceived impact of diagnosis and treatment [53].

Individuals who visited one the participating hospitals in the West Midlands between December 2005 and October 2011 with symptoms suspicious of bladder cancer were included. Patients with previous cancer of the urethra, bladder, ureter, or renal pelvis within the last decade were excluded. Inclusion criteria to participate in the BCPP cohort study were being over 18 years of age, able to provide consent and being fit for cystoscopy and/or surgical biopsy/resection. To collect data on information on social-demographic information, dietary intake, smoking behaviour and other health-related lifestyle behaviours, social support and quality of life and medical and supplement usage history, all participants were asked to fill in a semi-structured questionnaire under supervision of a research nurse and *Case Report Forms* (CRFs) were also collected by research nurses. The *European Organisation for the Research and Treatment of Cancer* (EORTC) QLQ-C30 v3 was used for the collection of the HRQoL data. This was repeated 3 months after diagnosis and hereafter yearly until 5 years after inclusion. Moreover, patients' medical records were continuously examined for 5 years after inclusion to check for recurrence or progression events and information on clinical treatment. After pathological assessment, some of the initially included participants (for whom baseline data

had already been collected by the research nurses) were not diagnosed with bladder cancer but with other non-malignant disease, such as bladder inflammation, so no follow-up data was collected for these participants. Results from this prospective cohort study are presented in **Chapter 2** (**Figure 1-3**).

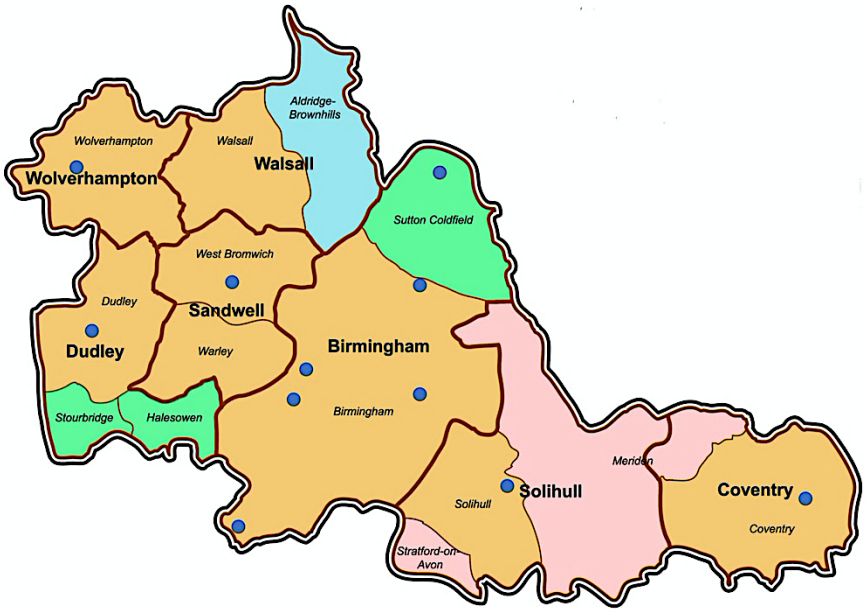


Figure 1-3 *Map of the Study BCPP Sites across the West-Midlands*
(Adapted from online source: https://www.webuyanystairlift.com/stairlift-removals-birmingham/)

1.5b-2 BLadder cancer Epidemiology and Nutritional Determinants (BLEND) consortium

The *BLadder cancer Epidemiology and Nutritional Determinants* (BLEND) consortium currently consists of 19 case-control studies and 16 cohort studies, including data from the already pooled *European Prospective Investigation into Cancer* and Nutrition (EPIC) [54-87]. Recruitment for the BLEND consortium is ongoing, but at this point there are 13,112 cases, 21,307 controls and 691,936 cohort members in the BLEND database. The aim of pooling these observational studies is to gain more knowledge on the impact of dietary patterns, nutrients and individual food items on bladder cancer risk [94].

Besides data on dietary intake and disease status, BLEND also provides data on smoking behaviour, age, sex, family history of cancer and tumour staging. After receiving databases from the participating studies, data was cleaned and recoded to the same codebook and

appended to form the BLEND database. Food items were coded according to the Eurocode 2 Core classification version 99/2, which contains coding for 2,362 food items [88].

The large sample size of this consortium allows for reliable estimation of relatively small effect sizes which are to be expected when investigating the impact of single food items or dietary patterns on (bladder) cancer risk which is the aim of the consortium; in addition, this large sample size also allows for the application of complex statistical models that need sufficient power to yield reliable results as described in **Chapter 3-7** (**Figure 1-4**).

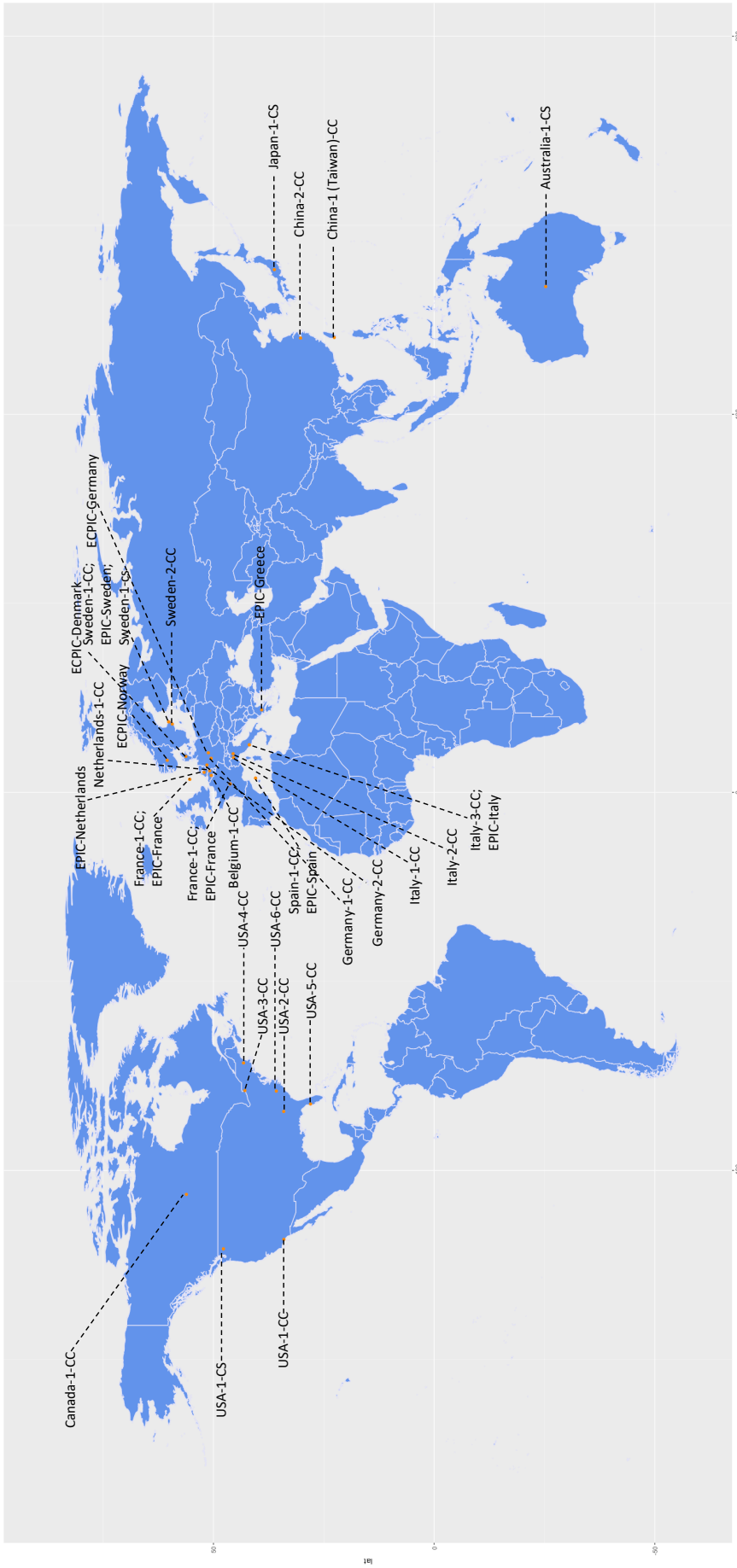


Figure 1-4 *Map of the Consortium BLEND Study Centres across the World*
Orange circle dots denote the study centres.
Abbreviations: CC, Case-Control study; CS, Cohort Study.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424. doi:10.3322/caac.21492.

2. Boyle P, Levin B. World cancer report 2008. IARC Press, International Agency for Research on Cancer; 2008 ISBN : 9789283204237; Record Number : 20103010665.

3. Continuous Update Project. Available at: https://wwwwcrforg/int/continuous-update-project. Accessed Dec 25th, 2019.

4. Waxman A. WHO global strategy on diet, physical activity and health. Food Nutr Bull. 2004;25(3):292-302. doi:10.1177/156482650402500310.

5. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018. doi:https://doi.org/10.3322/caac.21492.

6. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69(1):7-34. doi:10.3322/caac.21551.

7. Antoni S, Ferlay J, Soerjomataram I, Znaor A, Jemal A, Bray F. Bladder Cancer Incidence and Mortality: A Global Overview and Recent Trends. European urology. 2017;71(1):96-108. doi:10.1016/j.eururo.2016.06.010.

8. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA: a cancer journal for clinicians. 2020;70(1):7-30. doi:https://doi.org/10.3322/caac.21590.

9. Letašiová S, Medveďová A, Šovčíková A, Dušinská M, Volkovová K, Mosoiu C et al. Bladder cancer, a review of the environmental risk factors. Environmental Health. 2012;11(1):S11.

10. American Cancer Society: Key Statistics for Bladder Cancer 2019. Available at: https://wwwcancerorg/cancer/bladder-cancer/about/key-statisticshtml. Accessed Dec 25th, 2019.

11. Richters A, Aben KKH, Kiemeney L. The global burden of urinary bladder cancer: an update. World J Urol. 2019. doi:10.1007/s00345-019-02984-4.

12. Piyathilake C. Dietary factors associated with bladder cancer. Investigative and clinical urology. 2016;57 Suppl 1:S14-25. doi:10.4111/icu.2016.57.S1.S14.

13. Zeegers M. Diet and Fighting Bladder Cancer 1st Edition. Elsevier. 2019;ISBN: 9780128146781.

14. Woodburne RT. Anatomy of the bladder and bladder outlet. The Journal of urology. 1968;100(4):474-87. doi:https://doi.org/10.1016/S0022-5347(17)62556-4.

15. De Groat W. Anatomy and physiology of the lower urinary tract. The Urologic clinics of North America. 1993;20(3):383. doi:PMID: 8351765.

16. Marieb EN, Hoehn K. Urinary system. Human Anatomy and Physiology. 2006.

17. Lopez-Beltran A, Montironi R, Cheng L. Pathology of the Urinary Bladder. Cambridge University Press; 2016.

18. Kirkali Z, Chan T, Manoharan M, Algaba F, Busch C, Cheng L et al. Bladder cancer: epidemiology, staging and grading, and diagnosis. Urology. 2005;66(6):4-34. doi:10.1016/j.urology.2005.07.062.

19. Pashos CL, Botteman MF, Laskin BL, Redaelli A. Bladder cancer: epidemiology, diagnosis, and management. Cancer practice. 2002;10(6):311-22. doi:10.1046/j.1523-5394.2002.106011.x.

20. Van Der Molen AJ, Cowan NC, Mueller-Lisse UG, Nolte-Ernsting CC, Takahashi S, Cohan RH et al. CT urography: definition, indications and techniques. A guideline for clinical practice. European radiology. 2008;18(1):4. doi:10.1007/s00330-007-0792-x.

21. Furuse H, Ozono S. Transurethral resection of the bladder tumour (TURBT) for non-muscle invasive bladder cancer: Basic skills. International journal of urology. 2010;17(8):698-9. doi:10.1111/j.1442-2042.2010.02556.x.

22. Gospodarowicz MK, Brierley JD, Wittekind C. TNM classification of malignant tumours. John Wiley & Sons; 2017.

23. Knowles MA, Hurst CD. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. Nature Reviews Cancer. 2015;15(1):25. doi:10.1038/nrc3817.

24. Czerniak B. Molecular pathology and biomarkers of bladder cancer. Cancer Biomarkers. 2011;9(1-6):159-76. doi:10.3233/CBM-2011-0175.

25. Boustead GB, Fowler S, Swamy R, Kocklebergh R, Hounsome L, Section of Oncology B. Stage, grade and pathological characteristics of bladder cancer in the UK: B ritish Association of Urological Surgeons (BAUS) Urological Tumour Registry. BJU international. 2014;113(6):924-30. doi:10.1111/bju.12468.

26. Fernandez-Gomez J, Madero R, Solsona E, Unda M, Martinez-Piñeiro L, Gonzalez M et al. Predicting nonmuscle invasive bladder cancer recurrence and progression in patients treated

with bacillus Calmette-Guerin: the CUETO scoring model. The Journal of urology. 2009;182(5):2195-203. doi:10.1016/j.juro.2009.07.016.

27. Sylvester RJ, van der Meijden AP, Oosterlinck W, Witjes JA, Bouffioux C, Denis L et al. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. European urology. 2006;49(3):466-77. doi:10.1016/j.eururo.2005.12.031.

28. Sanli O, Dobruch J, Knowles MA, Burger M, Alemozaffar M, Nielsen ME et al. Bladder cancer. Nat Rev Dis Primers. 2017;3:17022. doi:10.1038/nrdp.2017.22.

29. American Society of Clinical Oncology: Bladder Cancer: Statistics. Available at: https://wwwcancernet/cancer-types/bladder-cancer/statistics. Accessed Dec 25th, 2019.

30. Millan-Rodriguez F, Chechile-Toniolo G, Salvador-Bayarri J, Palou J, Vicente-Rodriguez J. Multivariate analysis of the prognostic factors of primary superficial bladder cancer. The Journal of urology. 2000;163(1):73-8. doi:10.1016/s0022-5347(05)67975-x.

31. Ravvaz K, Walz ME, Weissert JA, Downs TM. Predicting nonmuscle invasive bladder cancer recurrence and progression in a United States population. The Journal of urology. 2017;198(4):824-31. doi:10.1016/j.juro.2017.04.077.

32. Prout GR, Marshall VF. The prognosis with untreated bladder tumors. Cancer. 1956;9(3):551-8. doi:10.1002/1097-0142(195605/06)9:3<551::aid-cncr2820090319>3.0.co;2-2.

33. Bray F, Ferlay J, Laversanne M, Brewster D, Gombe Mbalawa C, Kohler B et al. Cancer I ncidence in F ive C ontinents: Inclusion criteria, highlights from Volume X and the global status of cancer registration. International journal of cancer. 2015;137(9):2060-71. doi:10.1002/ijc.29670.

34. Anderson CB, Feurer ID, Large MC, Steinberg GD, Barocas DA, Cookson MS et al. Psychometric characteristics of a condition-specific, health-related quality-of-life survey: the FACT-Vanderbilt Cystectomy Index. Urology. 2012;80(1):77-83. doi:10.1016/j.urology.2012.01.090.

35. Hardt J, Filipas D, Hohenfellner R, Egle UT. Quality of life in patients with bladder carcinoma after cystectomy: first results of a prospective study. Quality of Life Research. 2000;9(1):1-12. doi:10.1023/a:1008927809013.

36. Ferriero M, Simone G, Rocchegiani A, Buscarini M, Papalia R, Alcini A et al. Early and late urodynamic assessment of Padua ileal bladder. Urology. 2009;73(6):1357-62. doi:10.1016/j.urology.2009.01.053.

37. Ferriero M, Simone G, Papalia R, Guaglianone S, Forastiere E, Gallucci M. Early and late urodynamic assessment of simplified Indiana pouch with multiple taeniamyotomies. BJU international. 2011;107(1):112-6. doi:10.1111/j.1464-410X.2010.09432.x.

38. Jerlström T, Andersson G, Carringer M. Functional outcome of orthotopic bladder substitution: a comparison between the S-shaped and U-shaped neobladder. Scandinavian journal of urology and nephrology. 2010;44(4):197-203. doi:10.3109/00365591003727577.

39. Månsson Å, Al Amin M, Malmström P-U, Wijkström H, Enein HA, Månsson W. Patient-assessed outcomes in Swedish and Egyptian men undergoing radical cystectomy and orthotopic bladder substitution—a prospective comparative study. Urology. 2007;70(6):1086-90. doi:10.1016/j.urology.2007.07.071.

40. Singh V, Yadav R, Sinha RJ, Gupta DK. Prospective comparison of quality-of-life outcomes between ileal conduit urinary diversion and orthotopic neobladder reconstruction after radical cystectomy: a statistical model. BJU international. 2014;113(5):726-32. doi:10.1111/bju.12440.

41. Zeegers MP, Tan FE, Dorant E, van den Brandt PA. The impact of characteristics of cigarette smoking on urinary tract cancer risk: a meta-analysis of epidemiologic studies. Cancer. 2000;89(3):630-9. doi:10.1002/1097-0142(20000801)89:3<630::aid-cncr19>3.3.co;2-h.

42. van Osch FH, Jochems SH, van Schooten F-J, Bryan RT, Zeegers MP. Quantified relations between exposure to tobacco smoking and bladder cancer risk: a meta-analysis of 89 observational studies. International journal of epidemiology. 2016;45(3):857-70. doi:10.1093/ije/dyw044.

43. Cumberbatch MG, Cox A, Teare D, Catto JW. Contemporary Occupational Carcinogen Exposure and Bladder Cancer: A Systematic Review and Meta-analysis. JAMA Oncol. 2015;1(9):1282-90. doi:10.1001/jamaoncol.2015.3209.

44. Al-Zalabani AH, Stewart KF, Wesselius A, Schols AM, Zeegers MP. Modifiable risk factors for the prevention of bladder cancer: a systematic review of meta-analyses. European journal of epidemiology. 2016;31(9):811-51. doi:10.1007/s10654-016-0138-6.

45. Cumberbatch MG, Windsor-Shellard B, Catto JW. The contemporary landscape of occupational bladder cancer within the United Kingdom: a meta-analysis of risks over the last 80 years. BJU international. 2017;119(1):100-9. doi:10.1111/bju.13561.

46. Research WCRFAIfC. Diet, Nutrition, Physical Activity and Cancer: A Global Perspective-The Third Expert Report. 2018.

47. ARC. Arsenic, metals, fibers and dusts. A review of human carcinogens. IARC monographs on the evaluation of carcinogenic risks to humans. 2012;100C. Available at: https://monographsiarcfr/wp-content/uploads/2018/06/ Accessed September 17, 2019.

48. Witlox WJA, van Osch FHM, Brinkman M, Jochems S, Goossens ME, Weiderpass E et al. An inverse association between the Mediterranean diet and bladder cancer risk: a pooled analysis of 13 cohort studies. European journal of nutrition. 2019. doi:10.1007/s00394-019-01907-8.

49. Bravi F, Spei ME, Polesel J, Di Maso M, Montella M, Ferraroni M et al. Mediterranean Diet and Bladder Cancer Risk in Italy. Nutrients. 2018;10(8). doi:10.3390/nu10081061.

50. Buckland G, Ros MM, Roswall N, Bueno-de-Mesquita HB, Travier N, Tjonneland A et al. Adherence to the Mediterranean diet and risk of bladder cancer in the EPIC cohort study. International journal of cancer. 2014;134(10):2504-11. doi:10.1002/ijc.28573.

51. Schulpen M, van den Brandt PA. Adherence to the Mediterranean Diet and Risks of Prostate and Bladder Cancer in the Netherlands Cohort Study. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2019;28(9):1480-8. doi:10.1158/1055-9965.Epi-19-0224.

52. Han J, Kamber M. Data mining concept and technology. *Publishing House of Mechanism Industry* Amsterdam Elsevier Inc. 2001;3rd:70-2.

53. Zeegers MP, Bryan RT, Langford C, Billingham L, Murray P, Deshmukh NS et al. The West Midlands Bladder cancer prognosis programme: rationale and design. BJU international. 2010;105(6):784-8. doi: https://doi.org/10.1111/j.1464-410X.2009.08849.x.

54. Bernstein L, Ross R. Cancer in Los Angeles County. Los Angeles, CA: University of Southern California. 1991.

55. Tang L, Zirpoli GR, Guru K, Moysich KB, Zhang Y, Ambrosone CB et al. Consumption of raw cruciferous vegetables is inversely associated with bladder cancer risk. Cancer Epidemiology and Prevention Biomarkers. 2008;17(4):938-44. doi:10.1158/1055-9965.EPI-07-2502.

56. Kellen E, Zeegers M, Lousbergh D, Paulussen A, Buntinx F. A Belgian case control study on bladder cancer: rationale and design. Archives of public health = Archives belges de sante publique. 2005;63:17-34.

57. Wakai K, Takashi M, Okamura K, Yuba H, Suzuki K-i, Murase T et al. Foods and nutrients in relation to bladder cancer risk: a case-control study in Aichi Prefecture, Central Japan. Nutrition and cancer. 2000;38(1):13-22. doi:https://doi.org/10.1207/S15327914NC381_3.

58. Lu C-M, Lan S-J, Lee Y-H, Huang J-K, Huang C-H, Hsieh C-C. Tea consumption: fluid intake and bladder cancer risk in Southern Taiwan. Urology. 1999;54(5):823-8. doi:https://doi.org/10.1016/S0090-4295(99)00281-2.

59. Pohlabeln H, Jöckel K-H, Bolm-Audorff U. Non-occupational risk factors for cancer of the lower urinary tract in Germany. European journal of epidemiology. 1999;15(5):411-9. doi:10.1023/a:1007595809278.

60. Steineck G, Hagman U, Gerhardsson M, Norell SE. Vitamin A supplements, fried foods, fat and urothelial cancer. A case-referent study in Stockholm in 1985–87. International journal of cancer. 1990;45(6):1006-11. doi:10.1002/ijc.2910450604.

61. METTLIN C, GRAHAM S. Dietary risk factors in human bladder cancer. American journal of epidemiology. 1979;110(3):255-63. doi:10.1093/oxfordjournals.aje.a112810.

62. Baena AV, Allam MF, Del Castillo AS, Díaz-Molina C, Tapia MJR, Abdel-Rahman AG et al. Urinary bladder cancer risk factors in men: a Spanish case–control study. European journal of cancer prevention. 2006;15(6):498-503. doi:10.1097/01.cej.0000215618.05757.04.

63. Brinkman MT, Karagas MR, Zens MS, Schned A, Reulen RC, Zeegers MP. Minerals and vitamins and the risk of bladder cancer: results from the New Hampshire Study. Cancer Causes & Control. 2010;21(4):609-19. doi:10.1007/s10552-009-9490-0.

64. La Vecchia C, Negri E, Decarli A, Benichou J. Attributable risks for bladder cancer in northern Italy. 1995. doi:10.1016/1047-2797(95)00057-7.

65. Shen M, Hung RJ, Brennan P, Malaveille C, Donato F, Placidi D et al. Polymorphisms of the DNA repair genes XRCC1, XRCC3, XPD, interaction with environmental exposures, and bladder cancer risk in a case-control study in northern Italy. Cancer Epidemiology and Prevention Biomarkers. 2003;12(11):1234-40. doi:PMID: 14652287.

66. Johnson K, Mao Y, Argo J, Dubois S, Semenciw R, Lava J. The National Enhanced Cancer Surveillance System: a case-control approach to environment-related cancer surveillance in Canada. Environmetrics: The official journal of the International Environmetrics Society.

1998;9(5):495-504. doi:https://doi.org/10.1002/(SICI)1099-095X(199809/10)9:5<495::AID-ENV318>3.0.CO;2-H.

67. Ovsiannikov D, Selinski S, Lehmann M-L, Blaszkewicz M, Moormann O, Haenel MW et al. Polymorphic enzymes, urinary bladder cancer risk, and structural change in the local industry. Journal of Toxicology and Environmental Health, Part A. 2012;75(8-10):557-65. doi:https://doi.org/10.1080/15287394.2012.675308.

68. Clavel J, Cordier S. Coffee consumption and bladder cancer risk. International journal of cancer. 1991;47(2):207-12. doi:https://doi.org/10.1002/ijc.2910470208.

69. Hemelt M, Hu Z, Zhong Z, Xie LP, Wong Y, Tam PC et al. Fluid intake and the risk of bladder cancer: Results from the South and East China case-control study on bladder cancer. International journal of cancer. 2010;127(3):638-45. doi:https://doi.org/10.1002/ijc.25084.

70. Cao W, Cai L, Rao JY, Pantuck A, Lu ML, Dalbagni G et al. Tobacco smoking, GSTP1 polymorphism, and bladder carcinoma. Cancer: Interdisciplinary International Journal of the American Cancer Society. 2005;104(11):2400-8. doi:https://doi.org/10.1002/cncr.21446.

71. Taylor JA, Umbach DM, Stephens E, Castranio T, Paulson D, Robertson C et al. The role of N-acetylation polymorphisms in smoking-associated bladder cancer: evidence of a gene-gene-exposure three-way interaction. Cancer research. 1998;58(16):3603-10. doi:PMID: 9721868.

72. Larsson SC, Andersson S-O, Johansson J-E, Wolk A. Fruit and vegetable consumption and risk of bladder cancer: a prospective cohort study. Cancer Epidemiology and Prevention Biomarkers. 2008;17(9):2519-22.

73. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol. 1997;26 Suppl 1:S6-14. doi:10.1093/ije/26.suppl_1.s6.

74. Tjonneland A, Olsen A, Boll K, Stripp C, Christensen J, Engholm G et al. Study design, exposure variables, and socioeconomic determinants of participation in Diet, Cancer and Health: a population-based prospective cohort study of 57,053 men and women in Denmark. Scandinavian journal of public health. 2007;35(4):432-41. doi:10.1080/14034940601047986.

75. Clavel-Chapelon F, van Liere MJ, Giubout C, Niravong MY, Goulard H, Le Corre C et al. E3N, a French cohort study on cancer risk factors. E3N Group. Etude Epidemiologique aupres de femmes de l'Education Nationale. European journal of cancer prevention : the official journal

of the European Cancer Prevention Organisation (ECP). 1997;6(5):473-8. doi:10.1097/00008469-199710000-00007.

76. Boeing H, Korfmann A, Bergmann MM. Recruitment procedures of EPIC-Germany. European Investigation into Cancer and Nutrition. Annals of nutrition & metabolism. 1999;43(4):205-15. doi:10.1159/000012787.

77. 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. doi:10.1079/phn2002394.

78. Panico S, Dello Iacovo R, Celentano E, Galasso R, Muti P, Salvatore M et al. Progetto ATENA, a study on the etiology of major chronic diseases in women: design, rationale and objectives. European journal of epidemiology. 1992;8(4):601-8.

79. Manjer J, Carlsson S, Elmstahl S, Gullberg B, Janzon L, Lindstrom M et al. The Malmo Diet and Cancer Study: representativity, cancer incidence and mortality in participants and non-participants. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP). 2001;10(6):489-99. doi:10.1097/00008469-200112000-00003.

80. 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. doi:10.1080/14034950310001432.

81. Beulens JW, Monninkhof EM, Verschuren WM, van der Schouw YT, Smit J, Ocke MC et al. Cohort profile: the EPIC-NL study. Int J Epidemiol. 2010;39(5):1170-8. doi:10.1093/ije/dyp217.

82. Davey GK, Spencer EA, Appleby PN, Allen NE, Knox KH, Key TJ. EPIC-Oxford: lifestyle characteristics and nutrient intakes in a cohort of 33 883 meat-eaters and 31 546 non meat-eaters in the UK. Public health nutrition. 2003;6(3):259-69. doi:10.1079/phn2002430.

83. Day N, Oakes S, Luben R, Khaw KT, Bingham S, Welch A et al. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. British journal of cancer. 1999;80 Suppl 1:95-103.

84. Lund E, Dumeaux V, Braaten T, Hjartaker A, Engeset D, Skeie G et al. Cohort profile: The Norwegian Women and Cancer Study--NOWAC--Kvinner og kreft. Int J Epidemiol. 2008;37(1):36-41. doi:10.1093/ije/dym137.

85. van den Brandt PA, Goldbohm RA, van 't Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. Journal of clinical epidemiology. 1990;43(3):285-95. doi:10.1016/0895-4356(90)90009-E.

86. White E, Patterson RE, Kristal AR, Thornquist M, King I, Shattuck AL et al. VITamins And Lifestyle cohort study: study design and characteristics of supplement users. Am J Epidemiol. 2004;159(1):83-93. doi:10.1093/aje/kwh010.

87. Ozasa K, Shimizu Y, Sakata R, Sugiyama H, Grant EJ, Soda M et al. Risk of cancer and non-cancer diseases in the atomic bomb survivors. Radiation protection dosimetry. 2011;146(1-3):272-5. doi:10.1093/rpd/ncr168.

88. Poortvliet E, Klensin J, Kohlmeier L. Rationale document for the Eurocode 2 food coding system (version 91/2). *Eur J Clin Nutr* 1992;46(Suppl 5):S9-S24.

Chapter 2

Health-Related Quality of Life in Bladder Cancer Patients around the Time of Diagnosis

Evan Y.W. Yu, Duncan Nekeman, Lucinda J. Billingham, Nicholas D. James, KK Cheng, Richard T. Bryan, Anke Wesselius and Maurice P. Zeegers

Adapted from BJU INT 2019 Jun; 124 (6): 984-991

Abstract

Objective: To quantify the *health-related quality of life* (HRQoL) of bladder cancer patients around the time of diagnosis and to test the hypotheses of a 2-factors model for the QLQ-C30.

Methods: Within the *Bladder Cancer Prognoses Programme* (BCPP), a multi-centre cohort study, sociodemographic data were collected using semi-structured face-to-face interviews. Answers to the QLQ-C30 were transformed into a scale from 0 to 100. HRQoL data was analysed in multivariate analyses. The hypothesized two-factor (Physical and Mental Health) domain structure of the QLQ-C30 was also tested with *confirmatory factor analysis* (CFA).

Results: Totally, 1,160 (78%) participants completed the questionnaire after initial visual diagnosis and before pathological confirmation. Despite non-muscle invasive bladder cancer (NMIBC) having a higher HRQoL than muscle invasive bladder cancer (MIBC), only the domain Role Functioning was clinically significantly better in NMIBC patients. Age, sex, bladder cancer stage and comorbidity all had a significant influence on QLQ-C30 scores. CFA showed an overall good fit of the hypothesized 2-factors model.

Conclusion: This study has identified a baseline reference value for HRQoL for bladder cancer patients, which allows for better evaluation of any changes in HRQoL as disease progresses or after treatment. In addition, a 2-factors (Physical and Mental Health) model was developed for the QLQ-C30.

Key Words: Bladder Cancer; Quality of Life; Cohort Study; QLQ-C30 Questionnaire

Introduction

Bladder cancer ranks as the 10th most frequently diagnosed cancer worldwide [1, 2]. In the UK, the disease accounts for approximately 10,300 new cases and 5,300 deaths per year [*Cancer Research UK* (CRUK) *cancer* statistics]; the majority of patients (75-80%) present with n*on-muscle-invasive bladder cancer* (NMIBC) [3]. Although not immediately life-threatening in the majority of cases, recurrence and progression of NMIBC remain significant issues, with up to 55% of patients experiencing recurrence within five years of diagnoses [4]. Current guidelines therefore recommend long-term surveillance [5, 6]. With the UK prevalence of NMIBC estimated at 46,500 by CRUK, at any one time there will be at least 40,000 patients requiring surveillance episodes at least once per year.

Typically, surveillance comprises of outpatient flexible cystoscopy with or without urine cytology [7, 8]. For patients with low-risk NMIBC, European Association of Urology (EAU) guidelines recommend follow-up cystoscopy and urine cytology at 3 and 12 months after transurethral resection of BC (TURBT), and then annually thereafter until 5 years. Patients with high-risk NMIBC undergo more intensive surveillance: every 3 months for the first 2 years, then every 6 months for the following 3 years, and annually thereafter, most likely for the rest of their lives [9]. If recurrence is detected, the tumour is resected and subsequent surveillance will start again, with the frequency determined by the risk category of the recurrence.

Each surveillance cystoscopy and urine cytology episode costs at least £ 533 [10]. As a result, bladder cancer is the most costly cancer to treat on a per patient basis from diagnosis to death [11, 12]. Furthermore, cystoscopy itself significantly increases the burden of disease, as it is an invasive procedure that causes pain and discomfort in about one-third of patients [13].

As survival amongst NMIBC patients is high compared with other cancers, and long-term follow-up is required, the disease can be considered as a chronic disease [14]. Therefore, *health-related quality of life* (HRQoL) plays an important role. However, to date, most HRQoL research has focused on *muscle-invasive bladder cancer* (MIBC) and, in particular, the differences between different urinary diversions after radical cystectomy [15-17].

Furthermore, these results are difficult to interpret, as there are often no baseline reference values for comparison. For MIBC patients, although HRQoL is affected more severely due to more radical treatments, a higher chance of metastasis, and worse overall prognosis, it is still an important measure. The few studies that have reported on HRQoL in NMIBC patients contradict each other [18-22]. Besides the sparse evidence, studies have often been hindered by small sample sizes, non-validated HRQoL questionnaires, and patient populations that are heterogeneous in terms of duration of follow-up, previous diagnoses of bladder cancer, and number of recurrences or progression.

Additionally, most studies do not take baseline HRQoL into account, which is important for the evaluation of any changes in a patient's HRQoL due to disease or treatment. Bladder cancer is a complex disease that often goes undetected for an extended period of time, especially in female patients [23]. This could imply that bladder cancer may have already affected patients' HRQoL at or around the time of diagnosis, and their HRQoL will not necessarily be the same as the general population. It is, therefore, important to study HRQoL at or around the time of bladder cancer diagnosis.

One of the reasons that HRQoL is not often taken into account in studies may result from the complex nature of the questionnaires used to measure HRQoL. Specifically, the commonly used HRQoL questionnaire QLQ-C30 [24] has 15 separate domains that need to be compared with each other in order to provide a complete assessment of patients' HRQoL. Higher order models would resolve the issue of too many individual domains. *Gundy et al.* (2011) [25] examined several of these models, including 1- and 2-factor models, and suggested that a 2-factor model divided into Physical and Mental health provided the best statistical fit.

The current study aims to be the first to quantify and compare HRQoL at or around the time of diagnoses in NMIBC and MIBC patients separately. Additionally, this study tests the hypotheses of a 2-factor model for QLQ-C30.

Methods

Study Design

The present study is part of the West Midlands' *Bladder Cancer Prognoses Programme* (BCPP), a multi-centre cohort study in the West Midlands, UK (ethics reference: 06/MRE04/65; clinicaltrials.gov registration number: NCT00553215). Details of the study have been published previously [26]. Briefly, adult patients (age ≥18 years) presenting at haematuria clinics in participating urology centres within the region were enrolled on the basis of abnormal cystoscopy findings suggestive of bladder cancer. Those who had a previous diagnosis of cancer of the urethra, bladder, ureter or renal pelvis within the last decade, HIV infection, or any other condition that might interfere with the safety of the participant were excluded. The enrolment period was from the 19th of December 2005 to the 21st of April 2011. All participants provided written informed consent before they were included in the study.

Data Collection

At the time of diagnoses, in some cases the research nurses were unable to speak to the patient face-to-face, in these cases they called the patient and filled out the questionnaire according to the answers of the patient. A consequence of this was that not all patients completed the HRQoL questionnaires at exactly the same time in their disease stage. on socio-demographics, health-related lifestyle, medical and drug history, dietary intake, social support and quality of life. The *European Organisation for the Research and Treatment of Cancer* (EORTC) QLQ-C30 v3 was used for the collection of the HRQoL data. The transformation of the scores is described in detail elsewhere [27]. The QLQ-C30 is a validated questionnaire specifically developed for measuring HRQoL in cancer patients. It has 30 items assessing 15 domains (1 global-HRQoL domain, 5 functional domains, and 9 symptom domains). Briefly, the answers are converted into a score from 0-100 where 100 is the best quality of life and 0 is the worst for all domains except the 9 symptom domains, for which 0 is the most problematic symptom and 100 is no symptoms at all. A difference of 10 points or more is considered clinically relevant [28].

The medical records of each patient were reviewed by trained research nurses, and clinicopathological characteristics of bladder cancers at diagnosis were prospectively gathered on dedicated *case report forms* (CRF). This comprised of pathological (pT) stage (according to the tumour, node, metastasis (TNM) 6th edition 2002 classification system [29]), grade (according to the *World Health Organization* (WHO) 1973 system [30]), size of the largest tumour, the number of visible tumours, and comorbidities. Where early re-resection (within 3 months after first surgery for bladder cancer) indicated an invasive tumour (≥pT2) contrary to the original assessment, then re-resection pT stage was recorded as the pT stage at diagnosis. Stage was recoded into a numerical outcome variable suitable for analyses by coding pTa tumours with a value "1" and all other stages in sequence. Comorbidities were defined as any pre-existing medical conditions, or conditions present at time of bladder cancer diagnosis. For this study, the number of comorbid conditions for each participant was recoded as 0, 1, 2, 3, 4, 5, >5. In addition, a separate variable was generated to recode stage as NMIBC (<pT2) and MIBC (pT2+).

Statistical Analyses

Means and *standard deviations* (SDs) were calculated for all domains of the QLQ-C30 to reflect reference values for NMIBC and MIBC patients. The effect of age, sex, stage and comorbidity on the 15 HRQoL domains from the QLQ-C30 were analysed using multivariate linear regression models. For this analysis, sex (male, female) and stage (pTis, pTa, pT1, pT2, pT3, pT4) and comorbidity (0, 1, 2, 3, 4, 5, >5) were included as categorical variables. Statistical differences between the HRQoL domains and NMIBC and MIBC were tested using the two-group mean-comparison (ANOVA). A factor analysis was used to detect the underlying hypothesized 2-factors model, a value of 0.45 was used as the cut-off value for factor loadings. We used *confirmatory factor analysis* (CFA) in order to test our hypothesis of a 2-factors model. The two overall HRQoL items were not included in the CFA, as this domain (overall HRQoL) is not considered to be part of HRQoL but rather a global domain.

Results

Socio-demographics

During the enrolment period, patients with symptoms related to bladder cancer accompanied by cystoscopically suspicious lesions were recruited. Out of 1,534 recruited participants, 1,183 were subsequently diagnosed with bladder cancer. Out of these, 1,160 had completed the questionnaire before they knew the histologically confirmed diagnosis and could therefore be included in the analyses for the current study.

The majority of patients were males (n=906; 78.1%), the average age of the patient population was 70.4 years (range=26-95 years). Over half of the patients were currently in a relationship (n=728; 62.8%). Most of the patients presented with NMIBC (n=890; 76.7%) with pTa as the most common tumour stage (n=575; 49.6%). In addition, 27.4% of the participants suffered from more than 5 comorbidities, as can be seen in **Table 2-1**.

HRQoL in patients NMIBC versus MIBC

The mean overall quality of life score was 69 (SD=23) in NMIBC patients, and 61 (SD=24) in MIBC patients. The functional domains for NMIBC patients ranged from 76 to 86, whilst in MIBC patients the same domains ranged from 72 to 81. In both groups (NMIBC and MIBC) the worst symptoms were Fatigue and Insomnia. In general, there were few differences in HRQoL between NMIBC and MIBC patients, with none of the scores differing by at least 10 points, except for Role Functioning (**Table 2-2**). It was notable that MIBC patients always scored lower than NMIBC patients on QLQ-C30, with statistically significant difference in seven domains (Global HRQoL, Physical Functioning, Role Functioning, Social Functioning, Fatigue, Pain, Dyspnoea, Insomnia, and Appetite Loss). The lowest scoring domain in both NMIBC and MIBC patients was the Global HRQoL (69 and 61 out of 100, respectively). Scores in all other domains were above 75 in the NMIBC group and above 65 in the MIBC group.

Table 2-1 *Socio-demographic Characteristics of Study Participants*

Demographic	Total	Sex		Age		Marital Status			Smoking Status			
		Male	Female	Mean (SD)	Range of age	With partner	Single	Missing	Never	Current	Former	Missing
No. of patients	1,160	906	254	70.4 (10.9)	26-95	728	344	88	217	228	612	103
%	100	78.1	21.9	—	—	62.8	29.7	7.5	18.7	19.7	52.8	8.8

Demographic	Tumour stage							Grade			
	pTis	pTa	pT1	pT2	pT3	pT4	Missing	Grade 1	Grade 2	Grade 3	Missing
No. of patients	14	575	301	233	1	17	19	254	336	540	30
%	1.2	49.6	25.9	20.1	0.1	1.5	1.6	21.9	29.0	46.5	2.6

Demographic	Subtype			Comorbidity						
	NMIBC	MIBC	Missing	0	1	2	3	4	5	>5
No. of patients	890	251	20	92	185	156	176	118	115	318
%	76.7	21.6	1.7	7.9	16.0	13.4	15.2	10.2	9.9	27.4

Missing values are not explicitly reported.

Abbreviations: SD=Standard Deviation; NMIBC, Non-Muscle Invasive Bladder Cancer; MIBC, Muscle Invasive Bladder Cancer.

Table 2-2 *HRQoL as Measured by the EORTC QLQ-C30 for NMIBC and MIBC Separately*

QLQ-C30[#]		**Global HRQoL**	***Functional domains (0 worst functioning; 100 best functioning)***				
			Physical Functioning	Role Functioning	Emotional Functioning	Cognitive Functioning	Social Functioning
NMIBC (*N*=890)	Mean	69	84	83	76	84	86
	SD	23	20	28	22	19	23
MIBC (*N*=251)	Mean	61	79	72	75	81	81
	SD	24	23	34	23	23	28
***P*-value**		<0.001	0.005	<0.001	0.536	0.036	0.003

		Symptom domains (0 worst symptoms; 100 no symptoms)								
		Dyspnoea	Fatigue	Nausea & Vomiting	Pain	Insomnia	Appetite Loss	Constipation	Diarrhoea	Financial Difficultie
NMIBC (*N*=890)	Mean	85	76	96	83	76	89	88	94	95
	SD	24	23	11	25	31	22	23	16	17
MIBC (*N*=251)	Mean	81	69	94	76	67	81	83	93	93
	SD	29	28	13	30	35	31	28	18	20
***P*-value**	Mean	0.015	<0.001	0.079	<0.001	<0.001	<0.001	0.012	0.471	0.147

P-value is for chi-square test.

[a] Range of QLQ-C30 is 0-100.

Abbreviations: HRQoL, Health Related Quality of Life; SD, standard deviation; EORTC, European Organization for Research and Treatment of Cancer; NMIBC, Non-Muscle Invasive Bladder Cancer; MIBC, Muscle Invasive Bladder Cancer.

From **Table 2-3**, it is clear that age had a statistically significant effect on nearly all Functional HRQoL domains, except for Social Functioning. Additionally, age had a statistically significant effect on Fatigue, Dyspnoea, Insomnia, Constipation, Diarrhea and Financial Difficulties. Increased age was detrimental in all statistically significant domains except for Financial Difficulties, Insomnia, Diarrhea and Emotional Functioning. The biggest effect was found in the Physical Functioning domain, where a 10-year increase in age would decrease the score by 4 points. Females had worse outcomes on every statistically significantly different HRQoL domain compared with males. Sex was the only variable that had a clinical impact (>10 points difference), for instance, on the domain Insomnia where males scored 12.19 points higher. All domains of HRQoL were negatively associated with the number of comorbidities. Patients with NMIBC scored better in every HRQoL domain of QLQ-C30 compared to MIBC patients, although this was not always a statistically significant difference, nor were any of the differences clinically significant.

Confirmatory factor analyses of 2-factors model

The factor analyses showed that 8 items (item 11, 13, 14, 15, 16, 17, 25, and 28) did not fit the 2-factors model (factor loading <0.45), most of which pertained to the 1-item domains such as Sleep (item 11), Appetite Loss (item 13), Constipation (item 16), Diarrhoea (item 17), and Financial Difficulties (item 28), as well as the 2-item domains such as, Nausea and Vomiting (item 14 and 15), and Cognitive Functioning (item 20 and 25). The domains Fatigue (item 10, 12, 18), Physical Functioning (item 1-5), Role Functioning (item 6 and 7), Dyspnoea (item 8), Pain (item 9 and 19), and Social Functioning (item 26 and 27) loaded onto factor 1 (Physical Health) and the domains Cognitive Functioning (item 20), and Emotional Functioning (item 21-24) loaded onto factor 2 (Mental Health). The sample size for the CFA, using the Physical/Mental Health model discussed above, was 1,036; 124 patients had missing data for at least one of the 28 items. The comparative fit index (CFI)=0.97, the Tucker-Lewis fit index (TLI)=0.95, and the *root mean squared error of approximation* (RMSEA)=0.05 (90% *confidence interval* (CI)=0.048-0.057, and *P*-close=0.20). In addition, the coefficient of determination (CD)=0.95 (interpreted as 95% of the variance in the observed variables being explained by the model) and the standardized root mean squared residual (SRMR)=0.034.

The differences between Physical/Mental Health stratified by the socio-demographic characteristics are shown in **Table 2-4**. Again, there were differences between the various categories within each stratum, amongst, differences between the most invasive tumour stage pT4 (n=17) and pT1 or below (n=890), as well as the most comorbidities (>5) and the non-comorbidity were larger than 10 points. Apart from tumour stage, the biggest difference could be found in the age strata where over 75s and under 65s differed by 8 points for Physical Health. The Physical and Mental health domains showed a good correlation (r=0.82).

Table 2-3 *Multivariate Linear Regression of Age, Sex, Stage and Comorbidity on each HRQoL Domain*

QLQ-C30		Global HRQoL	Physical F	Role F	Emotional F	Cognitive F	Social F	Fatigue	Nausea and vomiting
Age	β	-0.07	-0.42	-0.26	0.28	-0.09	0.01	-0.12	0.06
	P-value	0.015	<0.001	0.002	<0.001	0.024	0.836	0.029	0.072
Sex	β	-4.49	-4.83	-1.19	-6.99	0.61	-2.87	-5.26	-5.33
	P-value	0.007	0.001	0.577	<0.001	0.685	0.113	0.002	<0.001
Comorbidity	β	0.30	0.15	0.26	0.20	0.19	0.22	0.26	0.04
		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.027
Social Support	*P*-value	-6.85	-1.77	-8.16	-0.77	-1.78	-4.03	-5.02	-1.23
	β	<0.001	0.220	<0.001	0.640	0.236	0.026	0.004	0.166
Comorbidity	*P*-value	-2.83	-3.21	-3.26	-1.78	-1.82	-2.24	-3.61	-0.60
	β	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002

Table 2-3_continued *Multivariate Linear Regression of Age, Sex, Stage and Comorbidity on each HRQoL Domain*

QLQ-C30		Pain	Dyspnoea	Insomnia	Appetite loss	Constipation	Diarrhoea	Financial difficulties
Age	β	-0.02	-0.21	0.37	-0.02	-0.30	0.11	0.24
	P-value	0.824	0.004	<0.001	0.771	<0.001	0.020	<0.001
Sex	β	-4.01	0.06	-12.19	-7.81	-2.50	-2.72	0.07
	P-value	0.037	0.972	<0.001	<0.001	0.171	0.029	0.958
Comorbidity	β	0.25	0.09	0.25	0.25	0.13	0.11	0.20
		<0.001	0.021	<0.001	<0.001	0.002	<0.001	0.03
Social Support	*P*-value	-5.43	-2.64	-8.41	-7.06	-3.02	-0.67	-2.17
	β	0.005	0.153	<0.001	<0.001	0.098	0.592	0.097
Comorbidity	*P*-value	-3.56	-3.24	-3.22	-1.15	-1.28	-1.14	-1.05
	β	<0.001	<0.001	<0.001	0.003	0.001	<0.001	<0.001

β : beta, unstandardized coefficient of the multivariate linear regression analyses.
F represents a functional domain.
Abbreviation: HRQoL, Health related quality of life.

Table 2-4 *Physical and Mental Health Scores across the Different Socio-Demographic Variables*

Demographic		Sex		Age			Marital status		Smoking status		
		Male	Female	<=65 (1)	66-75(2)	>75 (3)	With partner	Single	Never	Current	Form
Physical Health	**Mean**	80***	75	83***	80	75	75***	81	82	78*	78
	SD	20	22	20	20	21	23	21	20	23	21
Mental Health	**Mean**	81*	78	81*	82	79	79	81	83	79	80
	SD	17	19	19	16	18	20	17	17	21	18

Demographic		Tumour Stage						Subtype	
		pTis	pTa	pT1	pT2	pT3	pT4	NMIBC	MIBC
Physical Health	**Mean**	82***	81	79	75	73	69	80***	74
	SD	21	19	21	23	—	32	20	23
Mental Health	**Mean**	85**	82	81	78	87	71	81*	78
	SD	13	16	18	18	—	25	17	19

Demographic		Comorbidity						
		0	1	2	3	4	5	>5
Physical Health	**Mean**	94***	93	88	82	83	79	70
	SD	14	13	17	19	20	22	25
Mental Health	**Mean**	87***	86	86	80	81	80	73
	SD	17	15	16	17	18	17	23

**=P-value <0.05; **=P-value <0.01; ***=P-value <0.001*

Abbreviations: SD, standard deviation; NMIBC, Non-Muscle Invasive Bladder Cancer; MIBC, Muscle Invasive Bladder Cancer.

Discussion

To our knowledge, this study is the first to investigate HRQoL of NMIBC and MIBC patients at this specific time point (*i.e.* around time of diagnosis). Interestingly, the HRQoL of both NMIBC and MIBC patients was relatively high: none of the mean values for each domain fell below 60, and patients in both groups even scored over 90 in a number of domains. In addition, we have identified that reduced HRQoL is common in patients with the features of aging, females, advanced tumor stage. Furthermore, bladder cancer cases with more comorbid conditions were much more likely to report poor HRQoL.

It is important to recognize that bladder cancer patients may have a different baseline HRQoL compared with the general population, as the disease may have already had an impact at or around the time of diagnosis. No reference data, however, is available for the general population in the UK; notwithstanding, the mean scores of HRQoL (QLQ-C30) for the present study population, are lower than data for the general populations of Germany, Norway, Sweden and Slovenia [31-34], but higher than the general population of the United States [35]. It is, therefore, possible that HRQoL reference values for bladder cancer patients depend on cultural factors; for example, Scandinavian populations have a high standard of living [36]. This could lead to a greater difference in HRQoL among bladder cancer patients from different countries.

Although it is more likely that cancer treatment (*i.e.* surgery, chemotherapy, radiation) will have an effect on HRQoL, an immediate change to an individual's HRQoL at diagnosis may depend on the type of cancer. Compared to UK cancer cases diagnosed with several other cancer types, including; breast [37], lung [38, 39], colorectal [40], head and neck [41], malignant mesothelioma [42], endometrial cancer [43], oral and oropharyngeal cancer [44], bladder cancer cases showed to have relatively higher HRQoL for all domains, which might be due to the poorer prognosis of these cancer types.

No clinically significant differences (>10 points) in HRQoL domains were found between NMIBC and MIBC patients, except for Role Functioning. This might be due to the unique time

point at which the questionnaire was administered, namely before histopathological confirmation of diagnosis. Previous studies have found significantly worse HRQoL values between 3 to 18 months after diagnosis [45-47]. However, in our study it is apparent that the disease has already affected patients such that a difference between NMIBC and MIBC patients can be identified in the Role Functioning domain (indicating that MIBC patients feel more restricted in fulfilling their role in society) even at this early time point in the patient pathway. Nevertheless, MIBC patients did not report higher pain levels or more restrictive Physical Functioning, possibly in contrast to what might be expected [48].

As demonstrated by multivariate analyses, age, sex, stage and comorbidity significantly influenced HRQoL. Overall, increased age, stage and comorbidity quantities were related to worse HRQoL. The tendency for females to have a lower HRQoL in bladder cancer cases is corresponded with found in the general populations [31, 49-51] and other cancer types [52, 53], and is in concordance with earlier population studies showing that, on average, females rated a higher proportion of health problems [54].

It was not surprising that age showed to be an important factor for defining HRQoL. Older participants showed a deterioration for most of the functional scales, except for emotional functioning which showed to slightly increased in older people, and social functioning which showed to be unrelated to age. These findings are in line with earlier studies in general populations from other European countries [31-34].

Comorbidity may in itself influence HQoL due to the unique physical impact of the specific health condition and may thus play an important role when assessing the HRQoL at time of bladder cancer diagnosis. We indeed observed that comorbidity relates to a decreased HRQoL, indicating that the presence of comorbidity before bladder cancer diagnosis may already have caused decrements in HRQoL. Since a more advanced tumour stage might cause more frightening of progression and death, it is not unexpected that tumour stage negatively affects HRQoL. Previous study in other cancer types (*e.g.* oropharyngeal cancer [55], prostate cancer

[56]) confirmed this finding. Therefore, both tumor-specific features and comorbidity should be evaluated in bladder cancer patients for better understanding of HRQoL.

It should be noted, however, that all of the shown statistical differences in our multivariate analysis did not meet the definition of clinical impact (all beta coefficients (β) <10), with the exception of female patients, who had substantially worse scores in the Insomnia domain. Furthermore, the beta coefficient (β) for age was relatively small per one-year difference, further implying limited clinical impact. The biggest impact of age can be found in Physical Functioning (β=0.56), although a 10-year age difference would only alter a patient's score by 5.6 points (on a scale from 0 to 100).

The factor analyses showed similar results to those of *Gundy et al* [25]: we detected a 2-factors model for the analyses of the QLQ-C30, even though not all the same items could be retained. We tested this model with CFA, which resulted in a well-fitting model. Interestingly, the items that did not load highly onto the factors were single item symptom domains such as, Insomnia, Appetite Loss, Constipation, Diarrhoea, and Financial Difficulties. This could suggest that they do not contribute as much to overall HRQoL in these patients. Having a 2-factors model indicating Physical and Mental Health would increase the usability of the HRQoL questionnaire QLQ-C30.

The EORTC has published a comprehensive report containing QLQ-C30 reference values for a wide variety of cancers [57, 58]. Compared with these reference values for cancer patients overall [59], BCPP participants had similar scores. EORTC reference scores for prostate cancer patients in stage I-II were comparable to NMIBC patients in BCPP (although prostate cancer patients did seem to have better Physical Functioning) and scores for MIBC patients in BCPP had comparable scores to stage III-IV prostate cancer patients (except for Physical Functioning, which was worse in prostate cancer patients). Reference values for lung cancer patients were worse than for MIBC patients in the BCPP cohort, with the exception of Cognitive Functioning, which was comparable.

As no UK QLQ-C30 data are available for bladder cancer patients at or around the time of diagnosis, this study aimed to provide the first data of this sort. In order to assess the generalisability of our study population and our findings, we compared our study demographics with the West Midlands Cancer Intelligence Unit data. We found no major differences with regard to age and sex (data not shown). In addition, the distribution of NMIBC (78%) and MIBC (22%) patients in BCPP corresponds with that in the literature from the UK [60]. Therefore, the results of this paper could be used as a reference population for future research into the HRQoL of bladder cancer patients at or around the time of diagnosis in the UK and beyond.

It is important to note that these data from BCPP have been captured from patients presenting in a conventional manner, having been referred from primary care to secondary care for the further investigation of symptoms, predominantly in a haematuria clinic setting. However, up to 18.5% of bladder cancer patients present as emergency cases [61], and it is likely that these patients will have different HRQoL than those presenting conventionally.

Conclusion

This is one of the largest bladder cancer patient cohort that has participated in HRQoL research. Additionally, this study captured the HRQoL at a unique time point, allowing for the identification of a baseline HRQoL measure. Our findings suggest that patients' HRQoL scores at or around time of bladder cancer diagnosis are almost comparable to the general population of Germany and possibly the UK. Only the domain Role Functioning was significantly better in NMIBC patients than MIBC patients at this timepoint.

The HRQoL scores given in this study are interesting for a clinical point of view. Comparisons with HRQoL scores at time of diagnosis give a better insight into any changes in a patient's HRQoL due to the disease or treatment.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018. doi:https://doi.org/10.3322/caac.21492.

2. Svatek RS, Hollenbeck BK, Holmang S, Lee R, Kim SP, Stenzl A et al. The economics of bladder cancer: costs and considerations of caring for this disease. European urology. 2014;66(2):253-62. doi:10.1016/j.eururo.2014.01.006.

3. Bryan RT, Zeegers MP, van Roekel EH, Bird D, Grant MR, Dunn JA et al. A comparison of patient and tumour characteristics in two UK bladder cancer cohorts separated by 20 years. BJU Int. 2013;112(2):169-75. doi:10.1111/bju.12032.

4. Sylvester RJ, van der Meijden APM, Oosterlinck W, Witjes JA, Bouffioux C, Denis L et al. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. European urology. 2006;49(3):466-77.

5. Roupret M, Babjuk M, Comperat E, Zigeuner R, Sylvester RJ, Burger M et al. European Association of Urology Guidelines on Upper Urinary Tract Urothelial Carcinoma: 2017 Update. European urology. 2018;73(1):111-22. doi:10.1016/j.eururo.2017.07.036.

6. Kaufman DS, Shipley WU, Feldman AS. Bladder cancer. The Lancet. 2009;374(9685):239-49.

7. Westhoff E, Wu X, Kiemeney LA, Lerner SP, Ye Y, Huang M et al. Dietary patterns and risk of recurrence and progression in non-muscle-invasive bladder cancer. International journal of cancer. 2018;142(9):1797-804. doi:10.1002/ijc.31214.

8. Moran ME, Moll FH. History of Cystoscopy. The History of Technologic Advancements in Urology. Springer; 2018. p. 3-20.

9. Babjuk M, Oosterlinck W, Sylvester R, Kaasinen E, Böhle A, Palou-Redorta J. EAU guidelines on non-muscle-invasive urothelial carcinoma of the bladder. European urology. 2008;54(2):303-14.

10. Mowatt G, Zhu S, Kilonzo M, Boachie C, Fraser C, Griffiths TRL et al. Systematic review of the clinical and cost-effectiveness of photodynamic diagnosis and urine biomarkers (FISH, ImmunoCyt, NMP22) and cytology for the detection and follow-up of bladder cancer. Health Technology Assessment. 2010;14(4).

11. Riley GF, Potosky AL, Lubitz JD, Kessler LG. Medicare payments from diagnosis to death for elderly cancer patients by stage at diagnosis. Medical care. 1995;33(8):828-41.

12. Svatek RS, Hollenbeck BK, Holmäng S, Lee R, Kim SP, Stenzl A et al. The economics of bladder cancer: costs and considerations of caring for this disease. European urology. 2014;66(2):253-62.

13. Koo K, Zubkoff L, Sirovich BE, Goodney PP, Robertson DJ, Seigne JD et al. The burden of cystoscopic bladder cancer surveillance: anxiety, discomfort, and patient preferences for decision making. Urology. 2017;108:122-8.

14. van Rhijn BW, Burger M, Lotan Y, Solsona E, Stief CG, Sylvester RJ et al. Recurrence and progression of disease in non–muscle-invasive bladder cancer: from epidemiology to treatment strategy. European urology. 2009;56(3):430-42.

15. Milowsky MI, Rumble RB, Lee CT. Guideline on muscle-invasive and metastatic bladder cancer (European Association of Urology guideline): American Society of Clinical Oncology clinical practice guideline endorsement summary. Journal of oncology practice. 2016;12(6):588-90.

16. Somani BK, MacLennan SJ, N'Dow J. Quality of life with urinary diversion. European Urology Supplements. 2010;9(10):763-71.

17. Botteman MF, Pashos CL, Hauser RS, Laskin BL, Redaelli A. Quality of life aspects of bladder cancer: a review of the literature. Quality of Life Research. 2003;12(6):675-88.

18. Gerharz EW, Månsson Å, Månsson W, editors. Quality of life in patients with bladder cancer2005: Elsevier.

19. Yoshimura K, Utsunomiya N, Ichioka K, Matsui Y, Terai A, Arai Y. Impact of superficial bladder cancer and transurethral resection on general health-related quality of life: an SF-36 survey. Urology. 2005;65(2):290-4.

20. Bohle A, Balck F, Von Wietersheim J, Jocham D. The quality of life during intravesical bacillus Calmette-Guerin therapy. The Journal of urology. 1996;155(4):1221-6.

21. Truta A, Popon TAH, Saraci G, Ghervan L, Pop IV. Health Related Quality of life in bladder cancer. Current approach and future perspectives. Clujul Medical. 2017;90(3):262.

22. Mason SJ, Downing A, Wright P, Hounsome L, Bottomley SE, Corner J et al. Health-related quality of life after treatment for bladder cancer in England. British journal of cancer. 2018;118(11):1518.

23. Bryan RT, Evans T, Dunn JA, Iqbal G, Bathers S, Collins SI et al. A Comparative Analysis of the Influence of Gender, Pathway Delays, and Risk Factor Exposures on the Long-term Outcomes of Bladder Cancer. European urology focus. 2015;1(1):82-9. doi:10.1016/j.euf.2015.01.001.

24. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. JNCI: Journal of the National Cancer Institute. 1993;85(5):365-76.

25. Gundy CM, Fayers PM, Groenvold M, Petersen MA, Scott NW, Sprangers MAG et al. Comparing higher order models for the EORTC QLQ-C30. Quality of Life Research. 2012;21(9):1607-17.

26. Zeegers MP, Bryan RT, Langford C, Billingham L, Murray P, Deshmukh NS et al. The West Midlands Bladder cancer prognosis programme: rationale and design. BJU international. 2010;105(6):784-8.

27. Fayers PM, Aaronson NK, Bjordal K, Groenveld M, Curran D, Bottomley A. The EORTC QLQ-C30 Scoring Manual Published by: European Organisation for Research and Treatment of Cancer. Brussels, Belgium. 2001;98.

28. King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. Quality of life research. 1996;5(6):555-67.

29. Gospodarowicz MK, Brierley JD, Wittekind C. TNM classification of malignant tumours. John Wiley & Sons; 2017.

30. Mostofi FK, Sobin LH, Torloni H, World Health O. Histological typing of urinary bladder tumours. 1973.

31. Schwarz R, Hinz A. Reference data for the quality of life questionnaire EORTC QLQ-C30 in the general German population. European Journal of Cancer. 2001;37(11):1345-51.

32. Hjermstad MJ, Fayers PM, Bjordal K, Kaasa S. Health-related quality of life in the general Norwegian population assessed by the European Organization for Research and Treatment of

Cancer Core Quality-of-Life Questionnaire: the QLQ=C30 (+ 3). Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1998;16(3):1188-96. doi:10.1200/jco.1998.16.3.1188.

33. Michelson H, Bolund C, Nilsson B, Brandberg Y. Health-related quality of life measured by the EORTC QLQ-C30: reference values from a large sample of the Swedish population. Acta Oncologica. 2000;39(4):477-84.

34. Velenik V, Secerov-Ermenc A, But-Hadzic J, Zadnik V. Health-related Quality of Life Assessed by the EORTC QLQ-C30 Questionnaire in the General Slovenian Population. Radiology and oncology. 2017;51(3):342-50. doi:10.1515/raon-2017-0021.

35. Smith AB, Jaeger B, Pinheiro LC, Edwards LJ, Tan HJ, Nielsen ME et al. Impact of bladder cancer on health-related quality of life. BJU Int. 2018;121(4):549-57. doi:10.1111/bju.14047.

36. Fayers PM, Machin D. Quality of life: the assessment, analysis and interpretation of patient-reported outcomes. John Wiley & Sons; 2013.

37. Hopwood P, Haviland J, Mills J, Sumo G, J MB. The impact of age and clinical factors on quality of life in early breast cancer: an analysis of 2208 women recruited to the UK START Trial (Standardisation of Breast Radiotherapy Trial). Breast (Edinburgh, Scotland). 2007;16(3):241-51. doi:10.1016/j.breast.2006.11.003.

38. Pompili C, Velikova G, White J, Callister M, Robson J, Dixon S et al. Poor preoperative patient-reported quality of life is associated with complications following pulmonary lobectomy for lung cancer. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery. 2017;51(3):526-31. doi:10.1093/ejcts/ezw363.

39. Nimako K, Ayite B, Priest K, Severn J, Fries HM, Gunapala R et al. A randomised assessment of the use of a quality of life questionnaire with or without intervention in patients attending a thoracic cancer clinic. European journal of cancer care. 2017;26(4). doi:10.1111/ecc.12402.

40. Radwan RW, Codd RJ, Wright M, Fitzsimmons D, Evans MD, Davies M et al. Quality-of-life outcomes following pelvic exenteration for primary rectal cancer. The British journal of surgery. 2015;102(12):1574-80. doi:10.1002/bjs.9916.

41. Llewellyn CD, McGurk M, Weinman J. Head and neck cancer: to what extent can psychological factors explain differences between health-related quality of life and individual

quality of life? The British journal of oral & maxillofacial surgery. 2006;44(5):351-7. doi:10.1016/j.bjoms.2005.06.033.

42. Arnold DT, Rowen D, Versteegh MM, Morley A, Hooper CE, Maskell NA. Testing mapping algorithms of the cancer-specific EORTC QLQ-C30 onto EQ-5D in malignant mesothelioma. Health and quality of life outcomes. 2015;13:6. doi:10.1186/s12955-014-0196-y.

43. Smits A, Lopes A, Das N, Bekkers R, Galaal K. The impact of BMI on quality of life in obese endometrial cancer survivors: does size matter? Gynecologic oncology. 2014;132(1):137-41. doi:10.1016/j.ygyno.2013.11.018.

44. Shepherd KL, Fisher SE. Prospective evaluation of quality of life in patients with oral and oropharyngeal cancer: from diagnosis to three months post-treatment. Oral oncology. 2004;40(7):751-7. doi:10.1016/j.oraloncology.2004.01.018.

45. Tyson MD, 2nd, Barocas DA. Quality of Life After Radical Cystectomy. The Urologic clinics of North America. 2018;45(2):249-56. doi:10.1016/j.ucl.2017.12.008.

46. Kulaksizoglu H, Toktas G, Kulaksizoglu IB, Aglamis E, Ünlüer E. When should quality of life be measured after radical cystectomy? European urology. 2002;42(4):350-5.

47. Singer S, Ziegler C, Schwalenberg T, Hinz A, Götze H, Schulte T. Quality of life in patients with muscle invasive and non-muscle invasive bladder cancer. Supportive Care in Cancer. 2013;21(5):1383-93.

48. Reeve BB, Potosky AL, Smith AW, Han PK, Hays RD, Davis WW et al. Impact of cancer on health-related quality of life of older Americans. JNCI: Journal of the National Cancer Institute. 2009;101(12):860-8.

49. Sullivan M, Karlsson J, Ware Jr JE. The Swedish SF-36 Health Survey—I. Evaluation of data quality, scaling assumptions, reliability and construct validity across general populations in Sweden. Social science & medicine. 1995;41(10):1349-58.

50. Brorsson B, Ifver J, Hays R. The Swedish health-related quality of life survey (SWED-QUAL). Quality of Life Research. 1993;2(1):33-45.

51. Hjermstad MJ, Fayers PM, Bjordal K, Kaasa S. Health-related quality of life in the general Norwegian population assessed by the European Organization for Research and Treatment of Cancer Core Quality-of-Life Questionnaire: the QLQ= C30 (+ 3). Journal of clinical oncology. 1998;16(3):1188-96.

52. Westby RP, Berg CJ, Leach C. Gender, race, BMI, and social support in relation to the health-related quality of life of cancer survivors: a report from the American Cancer Society's Study of Cancer Survivors II (SCS-II). Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2016;25(2):409-21. doi:10.1007/s11136-015-1084-6.

53. LeMasters T, Madhavan S, Sambamoorthi U, Kurian S. A population-based study comparing HRQoL among breast, prostate, and colorectal cancer survivors to propensity score matched controls, by cancer type, and gender. Psycho-oncology. 2013;22(10):2270-82. doi:10.1002/pon.3288.

54. Moum Tr. Self-assessed health among Norwegian adults. Social science & medicine. 1992;35(7):935-47.

55. Oates J, Davies S, Roydhouse JK, Fethney J, White K. The effect of cancer stage and treatment modality on quality of life in oropharyngeal cancer. The Laryngoscope. 2014;124(1):151-8. doi:10.1002/lary.24136.

56. Vanagas G, Mickeviciene A, Ulys A. Does quality of life of prostate cancer patients differ by stage and treatment? Scandinavian journal of public health. 2013;41(1):58-64. doi:10.1177/1403494812467503.

57. Scott NW, Fayers P, Aaronson NK, Bottomley A, de Graeff A, Groenvold M et al. EORTC QLQ-C30 reference values manual. 2008.

58. Sprangers MA, Bonnetain F. EORTC QLQ-C30. Encyclopedia of Quality of Life and Well-Being Research. 2014:1933-5.

59. Fayers P. Interpreting quality of life data: population-based reference data for the EORTC QLQ-C30. European journal of cancer. 2001;37(11):1331-4.

60. Boustead GB, Fowler S, Swamy R, Kocklebergh R, Hounsome L. Stage, grade and pathological characteristics of bladder cancer in the UK: British Association of Urological Surgeons (BAUS) urological tumour registry. BJU Int. 2014;113(6):924-30. doi:10.1111/bju.12468.

61. Abel GA, Shelton J, Johnson S, Elliss-Brookes L, Lyratzopoulos G. Cancer-specific variation in emergency presentation by sex, age and deprivation across 27 common and rarer cancers. British journal of cancer. 2015;112(s1):S129.

Chapter 3

The Association between Coffee Consumption and Bladder Cancer Risk: A Pooled Analysis from 13 Case-control Studies

Evan Y.W. Yu, Anke Wesselius, Frits van Osch, Mariana Carla Stern, Xuejuan Jiang, Eliane Kellen, Chih-Ming Lu, Hermann Pohlabeln, Gunnar Steineck, James Marshall, Md Farouk Allam, Carlo La Vecchia, Kenneth C. Johnson, Simone Benhamou, Zuo-Feng Zhang, Jerry Polesel, Jerry Polesel, Cristina Bosetti, Jack A. Taylor, Maurice P. Zeegers

Adapted from CANCER CAUSES & CONTROL 2019 May; 30 (8): 859-870

Abstract

Background: Inconsistent results for coffee consumption and bladder cancer risk have been shown in epidemiological studies. This research aims to increase the understanding of the association between coffee consumption and bladder cancer risk by bringing together worldwide case-control studies on this topic.

Methods: Data were collected from 13 case-control comprising of 5,911 cases and 16,172 controls. Pooled multivariate *odds ratios* (ORs), with corresponding 95% *confidence intervals* (CIs), were obtained using multilevel logistic regression models. Furthermore, linear dose-response relationships were examined using fractional polynomial models.

Results: No association of bladder cancer risk was observed with coffee consumption among smokers. However, after adjustment for age, sex and smoking, the risk was significantly increased for never smokers (ever versus never coffee consumers: OR_{model2}=1.30, 95% CI=1.06-1.59; heavy (>4 cups/day) coffee consumers versus never coffee consumers: OR_{model2}=1.52, 95% CI=1.18-1.97, P-trend=0.23). In addition, dose-response analyses, in both the overall population and among never smokers, also showed a significant increased bladder cancer risk for coffee consumption of more than 4 cups per day. Among smokers, a significant increased bladder cancer risk was shown only after consumption of more than 6 cups per day.

Conclusion: This research suggests that positive associations between coffee consumption and bladder cancer among never smokers but not smokers.

Key Words: Bladder Cancer; Coffee Consumption; Smoking; Dose-response Analyses; Population-attributable Risk

Introduction

Bladder cancer is the most common malignancy of the urinary tract and the seventh cause of death from cancer (2.8% of all cancer deaths), with nearly 430,000 new diagnoses and 165,000 deaths per year worldwide (2012) [1, 2]. Three-quarters of all bladder cancer cases occur in males [3], and most bladder cancer cases occur the United States, Canada, and the Europe [4-7]. As with many solid tumours, bladder cancer incidence increases with age and it rarely occurs before the age of 40-50 years [8]. Given that this cancer is easy to relapse, bladder cancer is reported to be the most expensive life-time treatment of all cancers raging from € 80,000 to € 160,000 per patient [9]. The strongest risk factors for bladder cancer occurrence, such as tobacco smoking and harmful chemicals [10], have long been identified. However, as the bladder is an excretory organ, the role of fluid consumption in the development of bladder cancer C could also be important.

Coffee is one of the most consumed beverages in the world. Since early 1970s, the possible relationship between coffee consumption and bladder cancer has been of considerable interest, when *Cole et al*. [11] suggested for the first time that coffee was a potential risk factor for bladder cancer. The *International Agency for Research on Cancer* (IARC) Monographs Programme in 1991, stated that there was limited evidence on the effect of coffee consumption on the bladder cancer risk and subsequently classified coffee as “possibly carcinogenic” (*group 2B; Monographs Volume 51*) [12]. Since then, several epidemiological studies focused on the relationship between coffee consumption and bladder cancer, however, results remained inconclusive [13-27]. In May 2016, a subsequent IARC Working Group of 23 scientists from 10 countries met to evaluate the carcinogenicity of drinking coffee and concluded that: “no consistent evidence of an association with drinking coffee, or of an exposure-response gradient”. This conclusion was based on evidence from 10 cohort studies and several population-based case-control studies conducted in Europe, the United States, and Japan [28]. An explanation for this inconsistency may be that previous studies on the relation between coffee consumption and bladder cancer risk lacked sufficient sample size to identify a significant association. In addition, since heavy coffee consumption is shown to be strongly associated with tobacco smoking [29], positive associations reported in some studies could possibly have been due to inadequate control for tobacco smoking. Moreover, several studies showed smoking to be interactive with caffeine in coffee [30-35], and, thereby, lead

to misleading results in un-stratified analysis on the relation between coffee consumption and bladder cancer risk.

The present study aims to update the understanding and find more conclusive answers on the influence of coffee consumption on the bladder cancer risk by bringing together available case-control studies on the topic including almost 6,000 bladder cancer cases.

Methods

Study Population

Data were derived from the *BLadder cancer Epidemiology and Nutritional Determinants* study (BLEND). BLEND is a large international epidemiology consortium aimed to pool data from available epidemiological studies on diet and bladder cancer. For the present study, 13 case-control studies (including 5,911 cases/16,172 controls), originated from 9 different countries in 3 continents (*i.e.* Europe, North America and Asia) had sufficient information on coffee consumption to be eligible for inclusion. Bladder cancer cases were diagnosed and histologically confirmed through each study centre of the included individual studies, with International Classification of Diseases (ICD) 9 or 10. Most of the bladder cancer cases were identified in 1990s.

Data Collection and Coding

Details on the methodology of the BLEND consortium have been described elsewhere [36]. Taking into account the local context of the included studies, different dietary assessment methods were adopted: 1) self-administrated *food frequency questionnaire* (FFQ) were used in Germany-1 [37], USA-2 [38], Canada-1 [39] , France-1 [40], USA-3 [41], USA-4 [42]; 2) FFQ administered by a trained interviewer were used in USA-1 [43], Beigium-1 [44], China-1 [45], Sweden-1 [46], Spain-1 [47], Italy-1 [48], Italy-2 [49]. Coffee consumption was categorized using the hierarchal Eurocode 2 food coding system developed by the European Union [50]. To obtain unified consumption across studies, weekly, monthly or yearly coffee cups were converted to daily cups of coffee consumption.

In addition to information on coffee and other dietary intake data, the BLEND dataset also included data on; study characteristics (design, method of dietary assessment, recall time of

dietary consumption and geographical region), participant demographics (age and sex), smoking status and smoking pack-years.

Statistical Analyses

To assess the influence of coffee consumption on the bladder cancer risk, multilevel logistic regression analyses were used to estimate the pooled *odds ratios* (ORs) and 95% *confidence intervals* (CIs). Coffee consumption was expressed as: a) ever (individuals who drank coffee >0 cup/day) or never consumption; b) based on the available data, coffee consumption was divided into 6 categories: never, 0-1 cup/day, 1-2 cups/day, 2-3 cups/day, 3-4 cups/day, and more than 4 cups/day; c) caffeinated or decaffeinated coffee consumption. In addition, standardized analysis on coffee cup size were performed. For this we transformed a United States (U.S.) cup size to a 237ml coffee cup size according to U.S. Food and Drug Administration [51] and an Asian cup size to a 500 ml cup size according to the questionnaire used in the study of *Chih-Ming Lu et al.* [39].

The logistic regression models used never coffee consumers as the reference group and were computed as "crude model" (model 1), adjusted for age (years, continuous), sex (male and female), smoking (model 2), or fully adjustment (model 3) additionally adjusted other fluid consumption [*i.e.* water, liquid milk, tea, alcohol, carbonated drinks and juice; *millilitres* (ml)/day, continuous]. Smoking was defined as: 0 (never smokers); 1 [current light smokers (*i.e.* smoking less than 20 pack-years)]; 2 [current heavy smokers (*i.e.* smoking more than 20 pack-years)]; 3 [former light smokers (*i.e.* smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 4 [former heavy smokers (*i.e.* smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]. In addition, the effect of ever versus never coffee consumption was also assessed using a meta-analysis approach; for this, pooled ORs (PORs) and 95% CIs were calculated by using a random-effect model stratified by geographical regions (*i.e.* Europe, North America and Asia) and study designs (*i.e.* hospital-based case-control studies and population-based case-control studies). Due to the lack of data, the influence of caffeine on bladder cancer risk was only assessed by comparing ever caffeinated coffee consumers versus never coffee consumers, ever decaffeinated coffee consumers versus never coffee consumers, and ever caffeinated versus ever decaffeinated coffee consumers based on multilevel logistic regression (model 2). To understand the

relevance of the effect modification, the interaction terms between coffee consumption and age, sex, smoking status with pack-years were added to the model 2. P-interaction <0.10 was considered statistically significant where upon analyses were stratified for the covariate of interest to understand the relevance of the effect modification.

In our secondary analysis, a potential dose-response relationship between coffee consumption and bladder cancer was assessed by using fractional polynomial regression, in which the best fitting second order fractional polynomial regression model was defined as the model with the lowest deviance [52, 53]. A likelihood ratio test was used to assess the difference between the nonlinear and linear models to test for nonlinearity [54]. The results of the dose-response analyses were presented for each 1 coffee cup/day increment up to 10 cups/day with stratification by smoking status (*i.e.* ever smokers and never smokers) and (*i.e.* hospital-based case-control studies and population-based case-control studies). Adjustments (model 2) were made for age, sex and smoking (in overall population).

Finally, the *population-attributable risk* (PAR) of heavy coffee consumption (*i.e.* >4 cups/day) on bladder cancer risk was estimated for Europe and North America, using the pooled risk estimates and the proportion of bladder cancer incidence in the population of interest.

All statistical analyses were performed with *STATA* version 14 SE (Stata Corporation, Texas, USA). *P*-values below 0.05 were considered statistically significant.

Results

Baseline Characteristics

The baseline characteristics of the study population are shown in **Table 3-1**. Of more than 22,000 participants, 5,911 cases of bladder cancer (4,639 males, 1,272 females) were identified. The median age at baseline was 61.4 years for cases and 57.2 years for controls, respectively. Approximately 35% of participants reported drinking coffee more than 4 cups per day, with an average consumption of 4 cups/day overall. At baseline, a higher coffee consumption was observed among smokers (5 cups/day) compared to never smokers (3 cups/day). In addition, coffee consumption showed strong interaction with smoking status as

well as pack-years (P-interaction: 0.001 and 0.042, respectively), while not interaction was found with age (P-interaction: 0.17) and sex (P-interaction: 0.16).

Associations between Coffee Consumption and Bladder Cancer Risk

Ever versus never coffee consumption

The results comparing ever versus never coffee consumers are shown in **Table 3-2** and **Figure 3-1**. Overall, after adjustment for possible confounders, no statistically significant difference in bladder cancer risk could be observed between coffee consumers versus never coffee consumers (OR_{model2}=1.11, 95% CI=0.98-1.26; OR_{model3}=1.09, 95% CI=0.94-1.25). Among never smokers, a statistically significant association between coffee consumption and the risk of bladder cancer was found after further adjustment (OR_{model2}=1.30, 95% CI=1.06-1.59; OR_{model3}=1.31, 95% CI=1.03-1.66). For smokers, no significant association was observed comparing ever versus never coffee consumers. However, the estimates for former light smokers showed to be slightly higher than the estimates for other smokers. In addition, the meta-analysis stratified by geographical regions presented similar PORs based on model 2 ($POR_{Overall}$=1.11, 95% CI=0.96-1.25; POR_{Europe}=1.13, 95% CI=0.88-1.37; $POR_{North\text{-}America}$=1.11, 95% CI=0.88-1.34;) with low heterogeneity (I^2=0.0%; P=0.57).

Categories of coffee consumption with bladder cancer risk

The results of multilevel logistic regressions for subsequent categories of coffee consumption are shown in **Table 3-2**. Overall, coffee consumption of more than 4 cups/day results in an increased bladder cancer risk of 1.27 (95% CI=1.11-1.46, P-trend=0.05, model 2) compared to never coffee consumers. A similar increased risk was observed among never smokers when comparing high (>4 cups/day) coffee consumption to never coffee consumers (OR_{model2}=1.52, 95% CI=1.19-1.94, P-trend=0.23). Among smokers no significant association could be observed; however, former light smokers showed again slightly higher and borderline significant results compared to other type of smokers (OR_{model2}=1.41, 95% CI=0.99-2.02, P-trend=0.06). The coffee cup size standardized analysis also showed significantly increased bladder cancer risk estimates with more than 4 cups/day coffee consumption (**Supplementary Table 3-3**).

Results for the comparison of caffeinated and decaffeinated coffee are shown in **Supplementary Table 3-1 and Table 3-2.** For the analysis on caffeinated coffee consumption only 1 (France-1) out of the 13 included case-control studies could be included (including 187 cases/296 controls). For the analysis on decaffeinated coffee, two studies (Italy-1 and USA-4) had sufficient data to be included in our analyses (including 1,048 cases/1,487 controls). Consumers of caffeinated coffee showed significant increased risks (compared to decaffeinated coffee consumers: OR_{model2}=1.88, 95% CI=1.42-2.48; compared to never coffee consumers: OR_{model2}=1.52, 95% CI=1.06-2.21), whereas, decaffeinated coffee consumers showed a null association with bladder cancer compared to never coffee consumers.

Population-attributable Risks

Assuming an incidence of heavy coffee drinking of 28% and 17%, in Europe and North America respectively, a PAR of 7.94% for Europe and 4.45% for North America was observed.

Table 3-1 *Characteristics of the Study Population (5,911 Cases and 16,172 Controls) and Coffee Consumption Category According to Age, Sex and Smoking Status*

Characteristics		*N* (%)	Case (%)	Control (%)	Sex			
					Male (%)		Female (%)	
					Case	Control	Case	Control
Coffee Consumption (cups/day)	**Total**	22,083 (100)	5,911 (100)	16,172 (100)	4,639 (100)	8,983 (100)	1,272 (100)	7,189 (100)
	Never	2,372 (10.74)	499 (8.44)	1,873 (11.58)	369 (7.95)	1,002 (11.15)	130 (10.22)	871 (12.12)
	Ever	19,711 (89.26)	5,412 (91.56)	14,299 (88.42)	4,270 (92.05)	7,981 (88.85)	1,142 (89.78)	6,318 (87.88)
	≤ 1	2,752 (12.46)	659 (11.15)	2,093 (12.94)	488 (10.52)	1,240 (13.80)	171 (13.44)	853 (11.87)
	1-2	3,616 (16.37)	1,077 (18.22)	2,539 (15.70)	865 (18.65)	1,468 (16.34)	212 (16.67)	1,071 (14.90)
	2-3	2,962 (13.41)	716 (12.11)	2,246 (13.89)	560 (12.07)	1,240 (13.80)	156 (12.26)	1,006 (13.99)
	3-4	2,750 (12.45)	801 (13.55)	1,949 (12.05)	646 (13.93)	1,114 (12.40)	155 (12.19)	835 (11.61)
	> 4	7,631 (34.56)	2,159 (36.53)	5,472 (33.84)	1711 (36.88)	2,919 (32.49)	448 (35.22)	2,553 (35.51)
	Mean (±SD)	4.00 (±3.97)	4.58 (±4.84)	3.78 (±3.58)	4.70 (±4.99)	3.84 (±3.98)	4.13 (±4.21)	3.71 (±3.01)
	P [a]		<0.001		<0.001		0.06	
	P [b]							

Table 3-1_continued *Characteristics of the Study Population (5,911 Cases and 16,172 Controls) and Coffee Consumption Category According to Age, Sex and Smoking Status*

Characteristics		Age (Mean ± SD)		Smoking Status					
				Never (%)		Current (%)		Former (%)	
		Case	Control	Case	Control	Case	Control	Case	Control
Coffee Consumption (cups/day)	**Total**	5 911 (100)	16 172 (100)	1 103 (100)	6 186(100)	2 317 (100)	5 350 (100)	2 491 (100)	4 636 (100)
	Never	60.48 (±12.70)	59.92 (±8.49)	172 (15.59)	1 052 (17.01)	140 (6.04)	357 (6.67)	187 (7.51)	464 (10.01)
	Ever	5 412 (91.56)	14 299 (88.42)	931 (84.41)	5 134 (82.99)	2 177 (93.96)	4 993 (93.33)	2 304 (92.49)	4 172 (89.99)
	≤ 1	65.38 (±9.88)	61.18 (±8.52)	173 (15.68)	924 (14.94)	176 (7.60)	324 (6.06)	310 (12.44)	845 (18.23)
	1-2	64.36 (±10.36)	61.94 (±8.80)	248 (22.48)	1 120 (18.11)	417 (18.00)	782 (14.62)	412 (16.54)	637 (13.74)
	2-3	64.34 (±8.65)	61.57 (±8.15)	102 (9.25)	822 (13.29)	258 (11.14)	478 (8.93)	356 (14.29)	946 (20.41)
	3-4	61.61 (±9.98)	60.52 (±9.32)	144 (13.06)	783 (12.66)	359 (15.49)	751 (14.04)	298 (11.96)	415 (8.95)
	> 4	59.88 (±9.21)	58.92 (±10.13)	264 (23.93)	1 485 (24.01)	967 (41.74)	2 658 (49.68)	928 (37.25)	1 329 (28.67)
	Mean (±SD)	4.52 (±4.82)	3.80 (±3.59)	3.11 (±3.14)	3.01 (±2.84)	4.95 (±4.60)	4.92 (±3.75)	4.88 (±5.50)	3.49 (±3.92)
	P [a]	<0.001		<0.001		<0.001		<0.001	
	P [b]	0.17							

[a] Calculated by χ^2 -test for categorical variables and t-test for continuous variables.

Age at the time of recruitment.

Abbreviation: N, Number; SD, Standard Deviation.

P [a]: P-value <0.05 was considered statistically.

P [b]: P-interaction <0.10 was considered statistically significant.

Table 3-2 *Adjusted Odds Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Level Stratified by Smoking*

Subgroups		Overall (*N*=2,083)			Current Light Smokers (*N*=3,548)			Current Heavy Smokers (*N*=3,458)		
Models		Model 1[b]	Model 2[c]	Model 3[d]	Model 1[b]	Model 2[c]	Model 3[d]	Model 1[b]	Model 2[c]	Model 3[d]
Coffee Consumption (cup/day)	**Never**	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	Ever	1.42 (1.26-1.59)	1.11 (0.98-1.26)	1.09 (0.94-1.25)	1.01 (0.68-1.47)	0.97 (0.65-1.43)	1.16 (0.71-1.90)	1.01 (0.71-1.41)	1.01 (0.71-1.42)	0.92 (0.59-1.41)
	≤ 1	1.05 (0.91-1.22)	0.94 (0.80-1.10)	0.92 (0.77-1.10)	0.75 (0.43-1.30)	0.63 (0.36-1.11)	0.71 (0.37-1.35)	0.91 (0.60-1.39)	0.83 (0.54-1.28)	0.75 (0.45-1.24)
	1-2	1.35 (1.17-1.55)	1.12 (0.96-1.29)	1.04 (0.88-1.24)	0.99 (0.64-1.53)	0.92 (0.59-1.44)	1.09 (0.63-1.88)	1.08 (0.74-1.60)	0.99 (0.67-1.47)	0.78 (0.48-1.29)
	2-3	1.23 (1.06-1.42)	1.01 (0.86-1.17)	0.97 (0.81-1.15)	0.71 (0.42-1.19)	0.71 (0.42-1.21)	0.80 (0.43-1.49)	0.92 (0.63-1.35)	0.91 (0.62-1.35)	0.82 (0.51-1.32)
	3-4	1.51 (1.31-1.76)	1.16 (0.99-1.35)	1.15 (0.96-1.37)	1.31 (0.84-2.04)	1.24 (0.79-1.94)	1.46 (0.85-2.48)	0.97 (0.65-1.44)	0.97 (0.65-1.46)	0.89 (0.55-1.45)
	> 4	1.82 (1.60-2.07)	1.27 (1.11-1.46)	1.24 (1.06-1.45)	1.09 (0.72-1.63)	1.09 (0.72-1.66)	1.25 (0.78-2.07)	1.06 (0.74-1.51)	1.15 (0.80-1.66)	1.03 (0.66-1.61)
P-trend		0.04	0.05	0.11	0.34	0.23	0.14	0.66	0.57	0.34

Table 3-2_continued *Adjusted Odds Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Level Stratified by Smoking*

Subgroups		Former Light Smokers (*N*=3,314)			Former Heavy Smokers (*N*=3,202)			Never Smokers (*N*=7,289)		
Models		Model 1 [b]	Model 2 [c]	Model 3 [d]	Model 1 [b]	Model 2 [c]	Model 3 [d]	Model 1 [b]	Model 2 [c]	Model 3 [d]
Coffee Consumption (cup/day)	**Never**	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
	Ever	1.24 (0.91-1.70)	1.28 (0.93-1.75)	1.30 (0.90-1.88)	1.02 (0.76-1.37)	1.01 (0.75-1.35)	0.92 (0.66-1.27)	1.36 (1.12-1.67)	1.30 (1.06-1.59)	1.31 (1.03-1.66)
	≤ 1	1.03 (0.71-1.51)	1.01 (0.68-1.48)	1.05 (0.68-1.62)	0.83 (0.58-1.17)	0.80 (0.57-1.13)	0.72 (0.49-1.05)	1.31 (1.01-1.70)	1.18 (0.90-1.55)	1.28 (0.95-1.73)
	1-2	1.20 (0.82-1.76)	1.19 (0.81-1.75)	1.14 (0.74-1.77)	0.96 (0.67-1.37)	0.93 (0.65-1.33)	0.79 (0.53-1.18)	1.53 (1.20-1.96)	1.36 (1.05-1.76)	1.42 (1.06-1.89)
	2-3	1.36 (0.95-1.97)	1.44 (0.99-2.09)	1.50 (0.99-2.27)	0.90 (0.64-1.26)	0.89 (0.63-1.24)	0.78 (0.54-1.13)	1.04 (0.78-1.39)	0.96 (0.71-1.30)	0.97 (0.70-1.34)
	3-4	1.37 (0.90-2.08)	1.36 (0.89-2.08)	1.39 (0.87-2.21)	1.03 (0.71-1.49)	1.03 (0.71-1.49)	0.96 (0.64-1.44)	1.31 (0.99-1.74)	1.22 (0.91-1.64)	1.24 (0.91-1.70)
	> 4	1.29 (0.91-1.83)	1.41 (0.99-2.02)	1.38 (0.93-2.07)	1.28 (0.93-1.76)	1.28 (0.93-1.76)	1.20 (0.85-1.70)	1.52 (1.19-1.94)	1.52 (1.18-1.97)	1.51 (1.15-1.99)
P-trend		0.07	0.06	0.09	0.19	0.18	0.07	0.23	0.23	0.24

[c] Model 1: Crude model without adjustment.
[c, d] Model 2: Adjusted for age (years, continuous), sex (male or female) and smoking [if applicable, smoking was defined as: 0 (never smokers); 1 (current light smokers (i.e. smoking less than 20 pack-years)); 2 (current heavy smokers (i.e. smoking more than 20 pack-years)); 3 (former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)); 4 (former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)].
[c, d, e] Model 3: Additionally, adjusted for water consumption [millilitre (ml)/day, continuous], liquid milk consumption (ml/day, continuous), alcohol consumption (ml/day, continuous), carbonated drinks consumption (ml/day, continuous), tea consumption (ml/day, continuous) and juice consumption (ml/day, continuous).
Abbreviation: OR, odds ratio; CI, confidence interval; ml, millilitre.
Reference group was never coffee consumers.
P-trend <0.05 was considered statistically significant.

Dose-response Analyses

Dose-response relationships between coffee consumption and the risk of bladder cancer are displayed in **Figure 3-2**. The tests for nonlinearity were not statistically significant, hence, linear models were applied in the dose-response analyses. The curves for the overall population showed a slightly increased bladder cancer risks with the increment of coffee consumption. Similar results were found among ever and never smokers; however, among never smokers the increased bladder cancer risk was significant for consuming over 4 cups/day, while for ever smokers for consumption over 6 cups/day. Adjusted ORs and 95% CIs for 1 cup/day increment were 1.14 (95% CI: 1.05-1.24) in overall study population, 1.05 (95% CI: 1.02-1.15) in ever smokers, 1.16 (95% CI: 1.04-1.41) in never smokers. In addition, both curves for hospital-based studies and population-based studies showed increased bladder cancer risks with the increment of cups of coffee consumption per day. However, in hospital-based studies a statistically significant increase bladder cancer risk was observed for consuming over 2 cups/day, while for population-based studies a significant increase was observed for consuming over 5 cups/day.

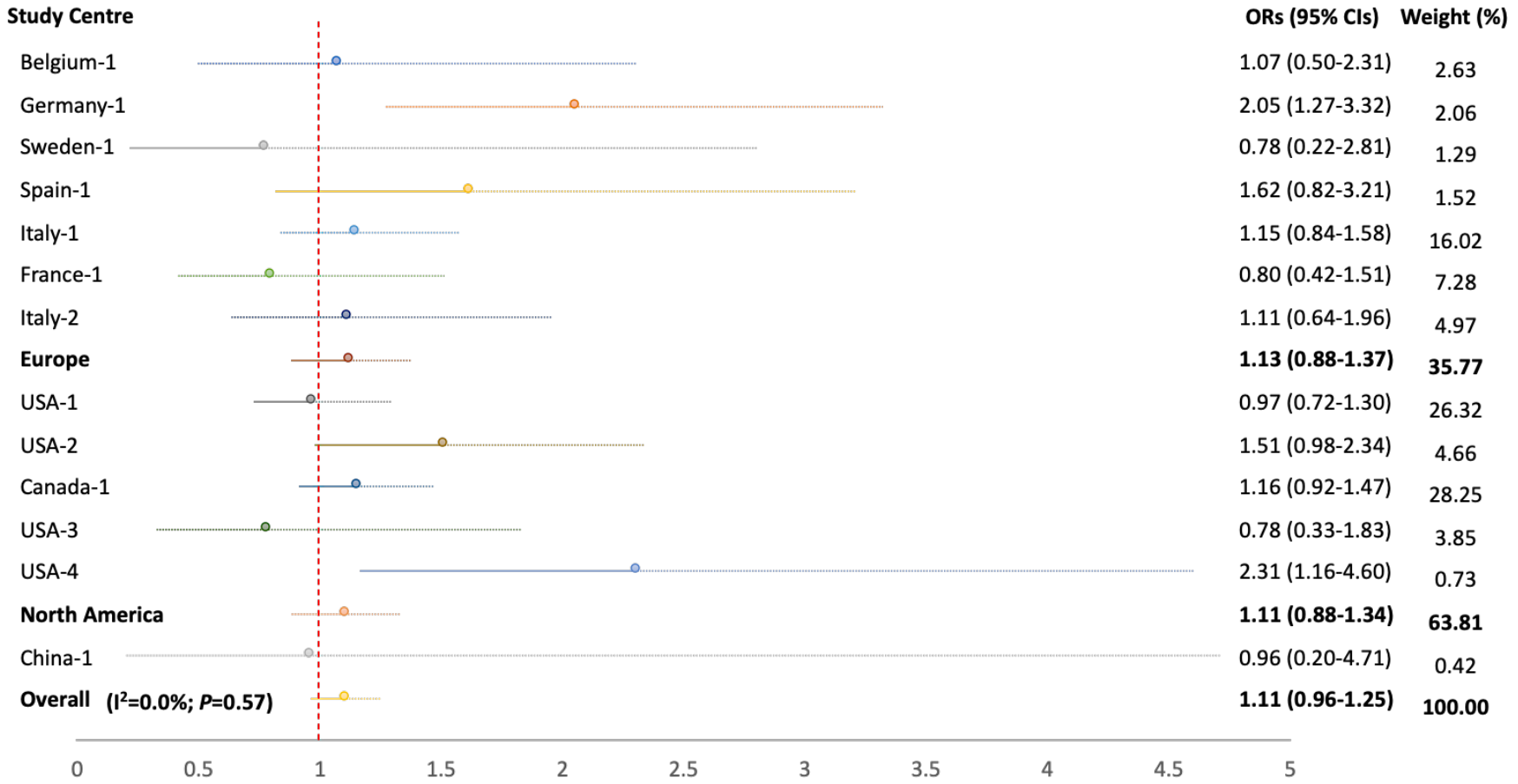


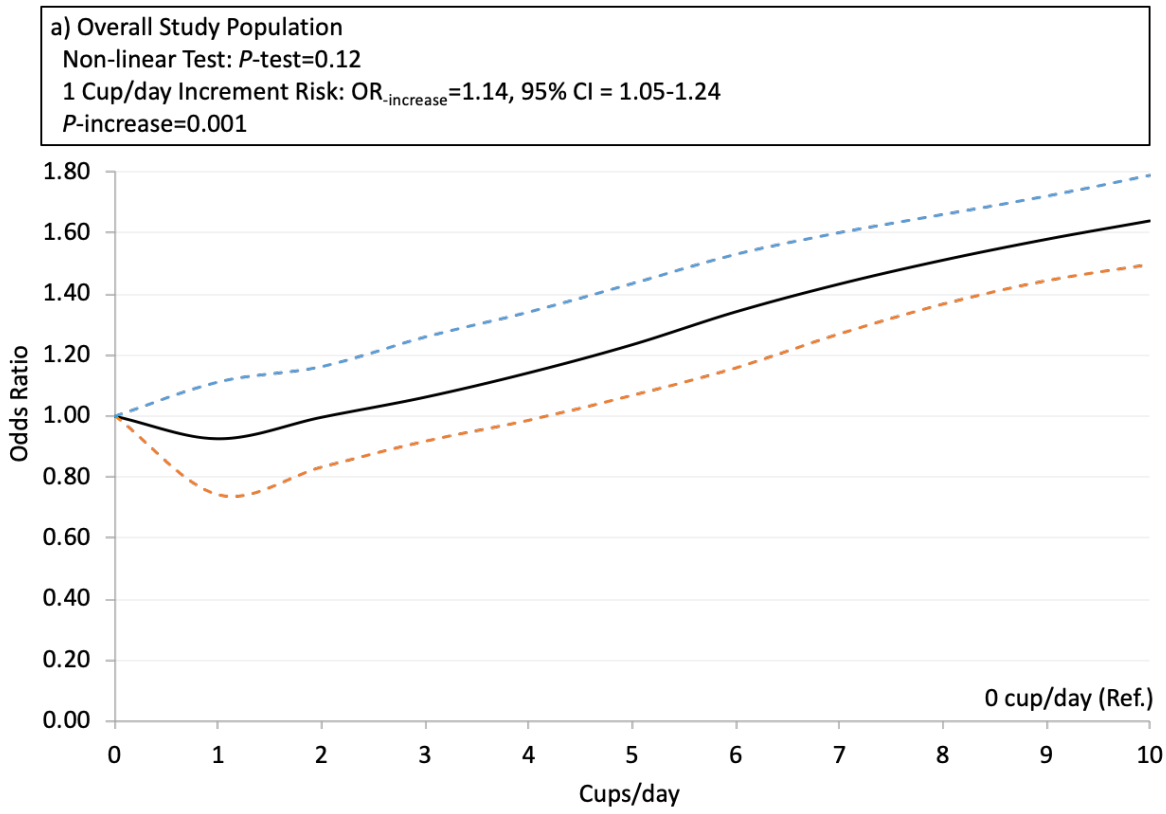
Figure 3-1 *Forest Plot of Meta-Analysis with ORs and 95% CIs for Ever and Never Coffee Consumption with Bladder Cancer Risk Adjusted for Age, Sex and Smoking by Geographical Regions*

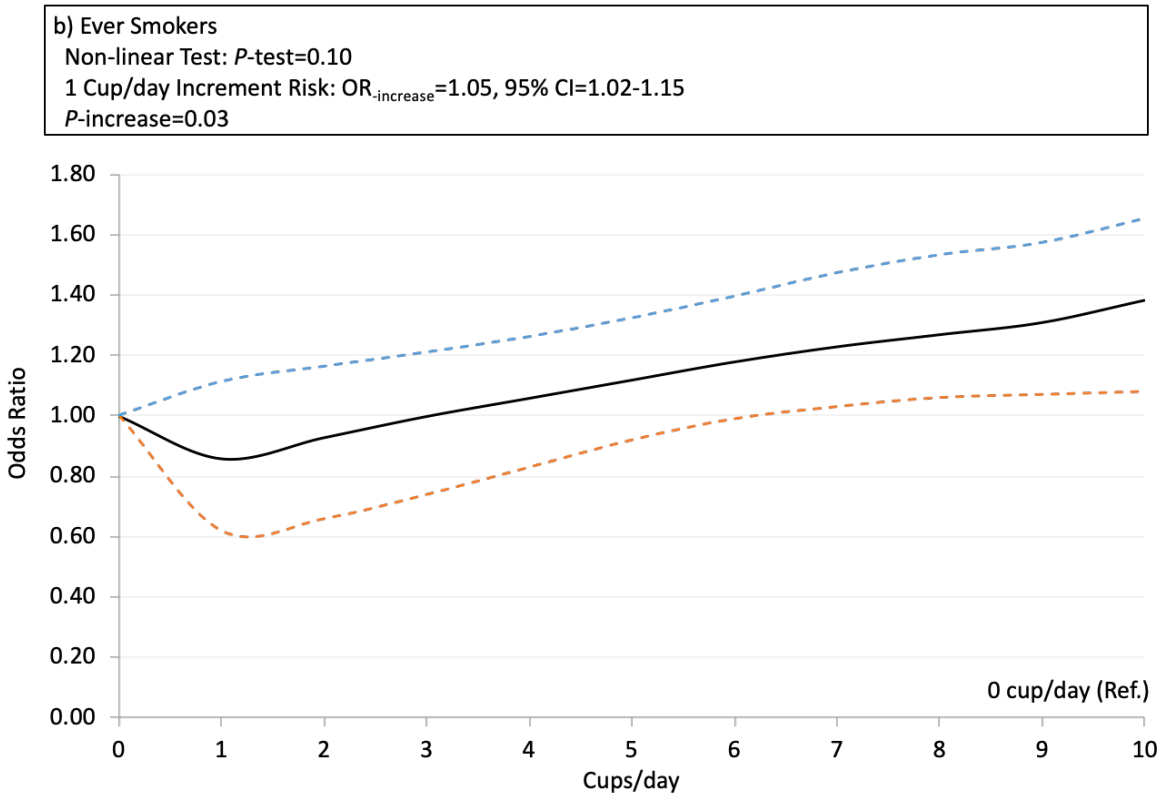
Abbreviation: OR, odds ratio; CI, confidence interval.

Circle dots denote the odds ratios (ORs); Horizontal lines represent the 95% confidence intervals (CIs); Weights are from random-effect model.

Europe*: pooled OR of studies from Europe;* ***North America****: pooled OR of studies from North America;* ***Overall****: pooled OR of all studies. Reference group was non coffee consumers.*

P-value <0.05 was considered statistically heterogeneous.


a) Overall Study Population
Non-linear Test: P-test=0.12
1 Cup/day Increment Risk: OR-increase=1.14, 95% CI = 1.05-1.24
P-increase=0.001
Odds Ratio
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
0 cup/day (Ref.)
0
1
2
3
4
5
6
7
8
9
10
Cups/day


b) Ever Smokers
Non-linear Test: P-test=0.10
1 Cup/day Increment Risk: OR-increase=1.05, 95% CI=1.02-1.15
P-increase=0.03
Odds Ratio
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
0 cup/day (Ref.)
0
1
2
3
4
5
6
7
8
9
10
Cups/day

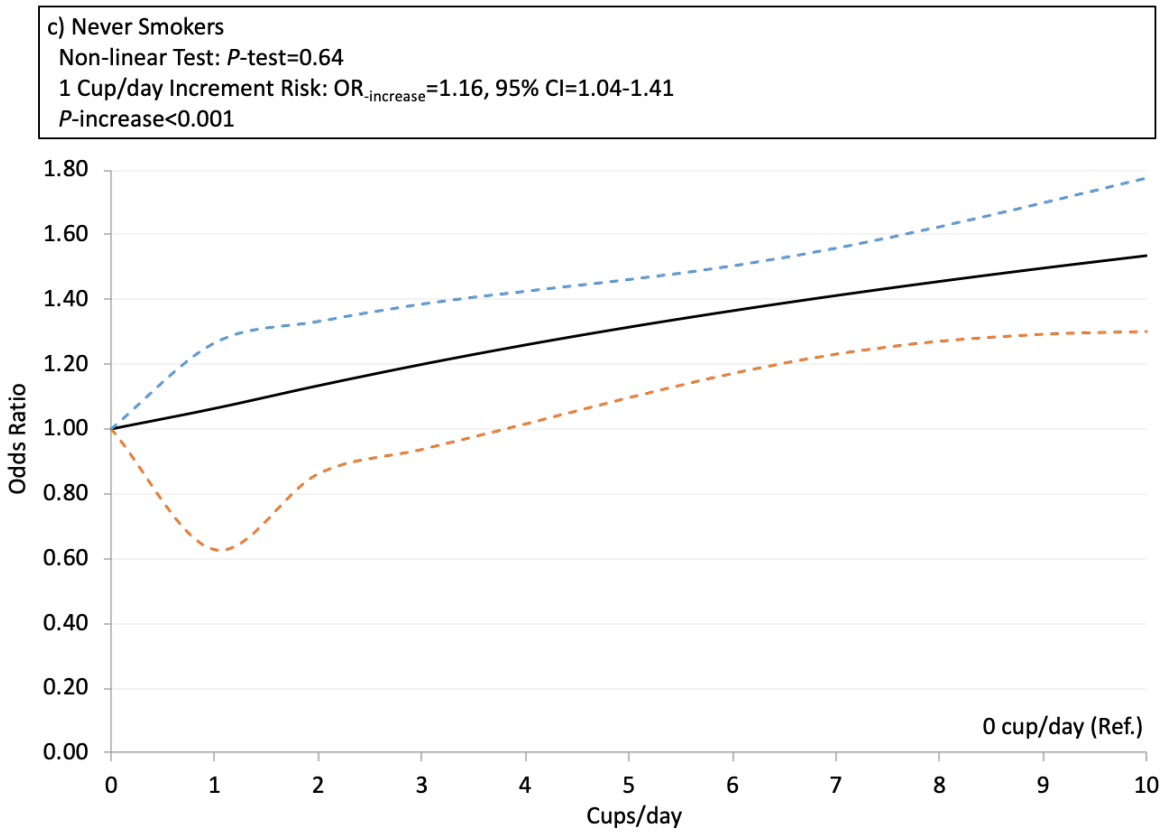


Figure 3-2 *Dose-response Relationships between Coffee Consumption and the Risk of Bladder Cancer among a) Overall study population; b) Ever Smokers; c) Never Smokers*
The solid lines represent the odds ratios (ORs). The dashed lines represent the 95% confidence intervals (CIs) for the trend. The ORs were adjusted for age, sex and smoking (in the overall study population) (model 2).
Ever Smokers: Current/Former Smokers.
Abbreviation: OR, odds ratio; CI, confidence interval.
Reference group was non coffee consumers.
P-test <0.05 was considered statistically significant.
P-increase <0.05 was considered statistically significant.

Discussion

This large multi centric study found an overall increased risk of bladder cancer with high (>4 cups/day) coffee consumption. In addition, we showed that this increased risk was only observed among never smokers, but not found among smokers.

The interaction between coffee and smoking has already been studied with some detail, and experimental studies showed that smokers eliminate caffeine faster, suggesting that the effect of coffee consumption on bladder cancer risk is lower among smokers [33, 55]. In fact,

faster metabolism of caffeine in smokers would allow them to consume higher levels before experiencing symptoms of caffeine toxicity [56-58]. Moreover, this hypothesis is strengthened by experimental studies reporting that the *cytochrome P450 1A2* (CYP1A2) metabolic pathway is upregulated by both caffeine and compounds in tobacco smoke, including nicotine and *polycyclic aromatic hydrocarbons* (PAHs) [59-61], so that the effect of caffeine is potentially weaker among smokers than among never smokers. These experimental studies are in line with our result showing an increased bladder cancer risk among never smokers only. Several epidemiological studies (both case-control and cohort) also suggested a higher bladder cancer risks for coffee consumers among never smokers only or a null association among smokers [40, 62-66]. A meta-analysis of epidemiological studies also suggested that the increased bladder cancer risk of coffee consumption was higher among never smokers than among smokers [31]. However, a more recent prospective cohort study (2017) [67] conducted in the United States, which showed that high coffee consumption (>4 cups/day) was positively associated with bladder cancer risk, could not observe an increased bladder cancer risk among never smokers only. This discrepancy may be due to the limited number of cases in their smoking-stratified analysis.

Coffee contains high content of caffeine, which has shown mutagenic effects in human cells [68], a proven influence on suppressing the activation of the protein kinase *ataxia-telangiectasia mutated* (ATM) and the phosphorylation of the kinase Chk2, both important for the activation of the tumour suppressor gene P53 [69, 70]. It could, therefore, be suggested that the increased bladder cancer risk is due to the caffeine content of coffee. Previous epidemiological studies already confirmed this hypothesis by showing a null association between bladder cancer risk and decaffeinated coffee [71] and an increased bladder cancer risk for caffeinated coffee [72, 73]. The present study also showed an increased bladder cancer risk when comparing caffeinated coffee consumers versus decaffeinated coffee consumers. However, due to lack of data, we were unable to perform further analyses, *i.e.* the association between the intensity of caffeinated coffee consumption and bladder cancer.

Besides caffeine, coffee also included other compounds, such as several phenolic compounds (*i.e.* chlorogenic, caffeic acid, ferulic, and coumaric acids), melanoidins, and diterpenes (*i.e.*

cafestol and kahweol), with anti-carcinogenic properties [74-79]. This might explain why the consumption of less than 4 coffee cups/day showed a null association with bladder cancer risk. In addition, experimental research on the mechanisms of action of coffee compounds on P53 suppression showed this effect to be concentration dependent [69, 80-83].

The present study found a 4 cups/day threshold for an increased bladder cancer risk in both the overall study population and among never smokers. This finding is in line with previous reported meta-analyses based on case-control studies, also showing an increased bladder cancer risk over 4 cups/day (respectively OR=1.29, 95% CI=1.12-1.48; OR=1.20, 95% CI=1.12-1.24) [84-86]. Among smokers, however, a threshold of 6 cups/day was observed, this again shows that the effect of coffee on bladder cancer risk might be influenced by the faster caffeine metabolism among smokers.

An issue that may arise in evaluating the influence of coffee consumption on bladder cancer risk in hospital-based case-control studies is that the controls includes patients suffering from a disease that may influence coffee intake (*i.e.* heart issues), resulting in inflated ORs. In the present study, therefore, stratified analyses by study design were performed (*i.e.* hospital-based controls or population-based controls), were performed (**Supplementary Figure 3-1**), and the association with coffee consumption was higher, though not significantly, in hospital-based studies than it was in population-based studies, where similar results were found (**Supplementary Figure 3-1** and **Supplementary Figure 3-2**).

For the present study, the estimates of PAR showed that in Europe 7.94% and in North America 4.45% of incidence of bladder cancer cases could be attributed to heavy coffee consumption (*i.e.* >4 cups/day). Unfortunately, the exact prevalence of heavy coffee drinking in the Eastern Asian population is yet unknown. However, since anecdotal evidence suggests that coffee consuming is on the rise in Eastern Asia, the PAR is expected to increase Eastern Asian countries over the next years.

Among the strengths of the *BLEND* study there is the large sample size, allowing to perform detailed analyses with enough statistical power to detect smaller effects, the study has also some limitations. First, it is known that the size of standard coffee cups is varied around the

world, and the effect of a cup of coffee on bladder cancer might, therefore, differ between different countries. However, our per-centre analyses as well our standardized analysis, in which we transformed a United States (U.S.) cup size to a 237ml coffee cup size according to U.S. Food and Drug Administration [51] and an Asian cup size to a 500 ml cup size according to the questionnaire used in the study of *Chih-Ming Lu et al.* [39], showed similar results (**Supplementary Table 3-3**: OR_{model2}=1.25, 95% CI=1.06-1.47, P-trend=0.18). Due to large heterogeneity among coffee cup size among European countries, Europe was not included in these standardized analyses. In addition, it is suggested that the strength of coffee brew may compensate for the different serving size between countries [87].

Second, it is often suggested that case-control studies are limited in showing causal relation, due to the potential recall bias of case-control studies. This might have led to a lower reliability of the results compared to those of cohort studies. However, although this issue has been addressed and analysed for its consequences in many epidemiological/methodological papers [88-91], no clear answer on the magnitude of the effect of this specific type of bias could be draw.

Thirdly, limited information was available on possible risk factors, other than age, sex and smoking, for the development of bladder cancer, such as *body mass index* (BMI), physical activity, *socioeconomic status* (SES), disinfection by-products, arsenic in the drinking water, and occupational exposures to potentially carcinogenic chemicals. Although, adjustments for these factors could have influenced the results, current literature shows that only a small proportion of bladder cancer cases can be attributed to these factors [72]. Last, although status as well as duration and intensity of smoking were taken into account in our analysis, the adjustment for smoking might still be imperfect due to differences in smoking practices (*e.g.* depth of inhalation or amount of inhalation), or differences in types of smoke exposure [67]. In addition, since smoking is perceived as a socially undesirable behaviour, the use of self-reported questionnaires for smoking status, duration and intensity might have led to underreporting of the actual smoking habits.

Conclusion

In summary, the present study, with more than 5,900 cases, observed an increased risk between high (>4 cups/day) coffee consumption and bladder cancer among never smokers, while no association with bladder cancer risk was observed with coffee consumption among smokers. Additionally, it indicates that around 7.94% of bladder cancer cases for Europe and 4.45% of bladder cancer cases for North America in the population might be attributable to heavy coffee consumption (>4 cups/day).

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer. 2015;136(5):E359-86. doi:10.1002/ijc.29210.

2. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. *CA Cancer J Clin* 2017;67(1):7-30. doi:10.3322/caac.21387.

3. Antoni S, Ferlay J, Soerjomataram I, Znaor A, Jemal A, Bray F. Bladder Cancer Incidence and Mortality: A Global Overview and Recent Trends. *Eur Urol* 2017;71(1):96-108. doi:10.1016/j.eururo.2016.06.010.

4. 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. doi:https://doi.org/10.3322/caac.20107.

5. Ploeg M, Aben KK, Kiemeney LA. The present and future burden of urinary bladder cancer in the world. World journal of urology. 2009;27(3):289-93. doi:10.1007/s00345-009-0383-3.

6. Marugame T, Mizuno S. Comparison of prostate cancer mortality in five countries: France, Italy, Japan, UK and USA from the WHO mortality database (1960–2000). Japanese Journal of Clinical Oncology. 2005;35(11):690-1. doi:https://doi.org/10.1093/jjco/hyi185.

7. Stewart B, Wild CP. World cancer report 2014. *International Agency for Research on Cancer, World Health Organization*: WHO Press, 2017.

8. Letašiová S, Medveďová A, Šovčíková A, Dušinská M, Volkovová K, Mosoiu C et al. Bladder cancer, a review of the environmental risk factors. Environmental Health. 2012;11(1):S11. doi:10.1186/1476-069X-11-S1-S11.

9. Botteman MF, Pashos CL, Redaelli A, Laskin B, Hauser R. The health economics of bladder cancer. Pharmacoeconomics. 2003;21(18):1315-30. doi:1170-7690/03/0018-1315/$30.00/0.

10. Johansson SL, Cohen SM, editors. Epidemiology and etiology of bladder cancer. Seminars in surgical oncology; 1997: Wiley Online Library.

11. Cole P. Coffee-drinking and cancer of the lower urinary tract. The Lancet. 1971;297(7713):1335-7. doi:https://doi.org/10.1016/S0140-6736(71)91890-3.

12. Cancer IAfRo. Coffee, tea, meat, methylxanthines and methylglyoxal. Apresentado em: IARC Working Group on the Evaluation of Carcinogenic Risks to Humans: Coffee, Lyon, 1990. 1990.

13. Pannelli F, La Rosa F, Saltalamacchia G, Vitali R, Petrinelli A, Mastrandrea V. Tobacco smoking, coffee, cocoa and tea consumption in relation to mortality from urinary bladder cancer in Italy. European journal of epidemiology. 1989;5(3):392-7. doi:10.1007/bf00144843.

14. RISCH HA, BURCH JD, MILLER AB, HILL GB, STEELE R, HOWE GR. Dietary factors and the incidence of cancer of the urinary bladder. American journal of epidemiology. 1988;127(6):1179-91. doi:10.1093/oxfordjournals.aje.a114911.

15. Ciccone G, Vineis P. Coffee drinking and bladder cancer. Cancer letters. 1988;41(1):45-52. doi:10.1016/0304-3835(88)90053-5.

16. Jensen O, Wahrendorf J, Knudsen J, Sørensen B. The Copenhagen case-control study of bladder cancer. II. Effect of coffee and other beverages. International journal of cancer. 1986;37(5):651-7. doi:10.1002/ijc.2910370503.

17. Bravo P, del Rey J, Sanchez J, Conde M. [Coffee and analgesics as risk factors for cancer of the bladder]. Arch Esp Urol. 1986;39(5):337-41. doi:PMID: 3753045.

18. Rebelakos A, Trichopoulos D, Tzonou A, Zavitsanos X, Velonakis E, Trichopoulos A. Tobacco smoking, coffee drinking, and occupation as risk factors for bladder cancer in Greece. Journal of the National Cancer Institute. 1985;75(3):455-61. doi:PMID: 3861898.

19. Gonzalez CA, Lopez-Abente G, Errezola M, Castejon J, Estrada A, Garcia-Mila M et al. Occupation, tobacco use, coffee, and bladder cancer in the county of Mataro (Spain). Cancer. 1985;55(9):2031-4. doi:10.1002/1097-0142(19850501)55:9<2031::aid-cncr2820550934>3.0.co;2-i.

20. Hopkins J. Coffee drinking and bladder cancer. Food Chem Toxicol. 1984;22(6):481-3. doi:10.1016/0278-6915(84)90336-3.

21. Marrett LD, Walter SD, Meigs JW. Coffee drinking and bladder cancer in Connecticut. Am J Epidemiol. 1983;117(2):113-27. doi:10.1093/oxfordjournals.aje.a113522.

22. Hartge P, Hoover R, West DW, Lyon JL. Coffee drinking and risk of bladder cancer. Journal of the National Cancer Institute. 1983;70(6):1021-6. doi:PMID: 6574270.

23. Howe GR, Burch JD, Miller AB, Cook GM, Esteve J, Morrison B et al. Tobacco use, occupation, coffee, various nutrients, and bladder cancer. Journal of the National Cancer Institute. 1980;64(4):701-13. doi:PMID: 6928984.

24. Abou-Daoud KT. Cancer of the bladder and cigarette smoking, coffee and alcohol drinking in Lebanon. J Med Liban. 1980;31(3):251-7. doi:PMID: 7218288.

25. Bross ID, Tidings J. Another look at coffee drinking and cancer of the urinary bladder. Prev Med. 1973;2(3):445-51. doi:10.1016/0091-7435(73)90040-6.

26. Zeitlin BR. Coffee and bladder cancer. Lancet (London, England). 1972;1(7759):1066. doi:10.1016/s0140-6736(72)91241-x.

27. Fraumeni JF, Jr., Scotto J, Dunham LJ. Coffee-drinking and bladder cancer. Lancet (London, England). 1971;2(7735):1204. doi:10.1016/s0140-6736(71)90523-x.

28. Loomis D, Guyton KZ, Grosse Y, Lauby-Secretan B, El Ghissassi F, Bouvard V et al. Carcinogenicity of drinking coffee, mate, and very hot beverages. The Lancet Oncology. 2016;17(7):877-8. doi:10.1016/s1470-2045(16)30239-x.

29. Bjorngaard JH, Nordestgaard AT, Taylor AE, Treur JL, Gabrielsen ME, Munafo MR et al. Heavier smoking increases coffee consumption: findings from a Mendelian randomization analysis. Int J Epidemiol. 2017;46(6):1958-67. doi:10.1093/ije/dyx147.

30. Swanson JA, Lee JW, Hopp JW. Caffeine and nicotine: a review of their joint use and possible interactive effects in tobacco withdrawal. Addict Behav. 1994;19(3):229-56. doi:10.1016/0306-4603(94)90027-2.

31. Zhou Y, Tian C, Jia C. A dose-response meta-analysis of coffee consumption and bladder cancer. Prev Med. 2012;55(1):14-22. doi:10.1016/j.ypmed.2012.04.020.

32. Benowitz NL, Peng M, Jacob P, 3rd. Effects of cigarette smoking and carbon monoxide on chlorzoxazone and caffeine metabolism. Clinical pharmacology and therapeutics. 2003;74(5):468-74. doi:10.1016/j.clpt.2003.07.001.

33. Zevin S, Benowitz NL. Drug interactions with tobacco smoking. An update. Clin Pharmacokinet. 1999;36(6):425-38. doi:10.2165/00003088-199936060-00004.

34. De Leon J, Diaz FJ, Rogers T, Browne D, Dinsmore L, Ghosheh OH et al. A pilot study of plasma caffeine concentrations in a US sample of smoker and nonsmoker volunteers. Progress in Neuro-Psychopharmacology and Biological Psychiatry. 2003;27(1):165-71. doi:https://doi.org/10.1016/S0278-5846(02)00348-2.

35. Plowchalk DR, Rowland Yeo K. Prediction of drug clearance in a smoking population: modeling the impact of variable cigarette consumption on the induction of CYP1A2. European journal of clinical pharmacology. 2012;68(6):951-60. doi:10.1007/s00228-011-1189-y.

36. Goossens ME, Isa F, Brinkman M, Mak D, Reulen R, Wesselius A et al. International pooled study on diet and bladder cancer: the bladder cancer, epidemiology and nutritional

determinants (BLEND) study: design and baseline characteristics. Archives of public health = Archives belges de sante publique. 2016;74:30. doi:10.1186/s13690-016-0140-1.

37. Pohlabeln H, Jockel KH, Bolm-Audorff U. Non-occupational risk factors for cancer of the lower urinary tract in Germany. European journal of epidemiology. 1999;15(5):411-9. doi:10.1023/a:1007595809278.

38. Mettlin C, Graham S. Dietary risk factors in human bladder cancer. Am J Epidemiol. 1979;110(3):255-63. doi:10.1093/oxfordjournals.aje.a112810.

39. Johnson K, Mao Y, Argo J, Dubois S, Semenciw R, Lava J. The National Enhanced Cancer Surveillance System: a case-control approach to environment-related cancer surveillance in Canada. Environmetrics: The official journal of the International Environmetrics Society. 1998;9(5):495-504. doi:https://doi.org/10.1002/(SICI)1099-095X(199809/10)9:5<495::AID-ENV318>3.0.CO;2-H.

40. Clavel J, Cordier S. Coffee consumption and bladder cancer risk. International journal of cancer. 1991;47(2):207-12. doi:https://doi.org/10.1002/ijc.2910470208.

41. Cao W, Cai L, Rao JY, Pantuck A, Lu ML, Dalbagni G et al. Tobacco smoking, GSTP1 polymorphism, and bladder carcinoma. Cancer. 2005;104(11):2400-8. doi:10.1002/cncr.21446.

42. Taylor JA, Umbach DM, Stephens E, Castranio T, Paulson D, Robertson C et al. The role of N-acetylation polymorphisms in smoking-associated bladder cancer: evidence of a gene-gene-exposure three-way interaction. Cancer Res. 1998;58(16):3603-10. doi:PMID: 9721868.

43. Jiang X, Castelao JE, Groshen S, Cortessis VK, Ross RK, Conti DV et al. Alcohol consumption and risk of bladder cancer in Los Angeles County. International journal of cancer. 2007;121(4):839-45. doi:10.1002/ijc.22743.

44. Kellen E, Zeegers M, Paulussen A, Van Dongen M, Buntinx F. Fruit consumption reduces the effect of smoking on bladder cancer risk. The Belgian case control study on bladder cancer. International journal of cancer. 2006;118(10):2572-8. doi:10.1002/ijc.21714.

45. Lu CM, Lan SJ, Lee YH, Huang JK, Huang CH, Hsieh CC. Tea consumption: fluid intake and bladder cancer risk in Southern Taiwan. Urology. 1999;54(5):823-8. doi:10.1016/s0090-4295(99)00281-2.

46. Steineck G, Hagman U, Gerhardsson M, Norell SE. Vitamin A supplements, fried foods, fat and urothelial cancer. A case-referent study in Stockholm in 1985-87. International journal of cancer. 1990;45(6):1006-11. doi:10.1002/ijc.2910450604.

47. Baena AV, Allam MF, Del Castillo AS, Diaz-Molina C, Requena Tapia MJ, Abdel-Rahman AG et al. Urinary bladder cancer risk factors in men: a Spanish case-control study. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP). 2006;15(6):498-503. doi:10.1097/01.cej.0000215618.05757.04.

48. Shen M, Hung RJ, Brennan P, Malaveille C, Donato F, Placidi D et al. Polymorphisms of the DNA repair genes XRCC1, XRCC3, XPD, interaction with environmental exposures, and bladder cancer risk in a case-control study in northern Italy. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2003;12(11 Pt 1):1234-40. doi:PMID: 14652287.

49. D'Avanzo B, La Vecchia C, Negri E, Decarli A, Benichou J. Attributable risks for bladder cancer in northern Italy. Annals of epidemiology. 1995;5(6):427-31. doi:10.1016/1047-2797(95)00057-7.

50. Poortvliet E, Klensin J, Kohlmeier L. Rationale document for the Eurocode 2 food coding system (version 91/2). *Eur J Clin Nutr* 1992;46(Suppl 5):S9-S24.

51. US Food Drug Administration. Guidance for industry: drug interaction studies—study design, data analysis, implications for dosing, and labeling recommendations. *Center for Drug Evaluation and Research* 2012:1-75.

52. Royston P, Altman DG. Approximating statistical functions by using fractional polynomial regression. *J Royal Stat Soc* 1997;46(3):411-422.

53. Jones BL, Nagin DS. A Stata plugin for estimating group-based trajectory models. *Sociological Methods & Research* 2012;42(4):608-613.

54. Bagnardi V, Zambon A, Quatto P, Corrao G. Flexible meta-regression functions for modeling aggregate dose-response data, with an application to alcohol and mortality. Am J Epidemiol. 2004;159(11):1077-86. doi:10.1093/aje/kwh142.

55. Mihi Y TK, Takahiko K, Koji M. Effects of lifestyle and genetic polymorphisms on consumption of coffee or black tea and urinary caffeine levels. *Biomarkers* 1998;3(4-5):367-377.

56. Joeres R, Klinker H, Heusler H, Epping J, Zilly W, Richter E. Influence of smoking on caffeine elimination in healthy volunteers and in patients with alcoholic liver cirrhosis. Hepatology. 1988;8(3):575-9. doi:10.1002/hep.1840080323.

57. Langmann P, Bienert A, Zilly M, Vath T, Richter E, Klinker H. Influence of smoking on cotinine and caffeine plasma levels in patients with alcoholic liver cirrhosis. Eur J Med Res. 2000;5(5):217-21. doi:PMID: 10806124.

58. Ossip DJ, Epstein LH. Relative effects of nicotine and coffee on cigarette smoking. Addict Behav. 1981;6(1):35-9. doi:10.1016/s0306-4603(81)80006-8.

59. Landi MT, Sinha R, Lang NP, Kadlubar FF. Human cytochrome P4501A2. IARC scientific publications. 1999(148):173-95. doi:PMID: 10493258.

60. Kalow W, Tang BK. Use of caffeine metabolite ratios to explore CYP1A2 and xanthine oxidase activities. Clinical pharmacology and therapeutics. 1991;50(5 Pt 1):508-19. doi:10.1038/clpt.1991.176.

61. Gunes A, Dahl ML. Variation in CYP1A2 activity and its clinical implications: influence of environmental factors and genetic polymorphisms. Pharmacogenomics. 2008;9(5):625-37. doi:10.2217/14622416.9.5.625.

62. Donato F, Boffetta P, Fazioli R, Aulenti V, Gelatti U, Porru S. Bladder cancer, tobacco smoking, coffee and alcohol drinking in Brescia, northern Italy. European journal of epidemiology. 1997;13(7):795-800. doi:10.1023/a:1007453322899.

63. Escolar Pujolar A, Gonzalez CA, Lopez-Abente G, Errezola M, Izarzugaza I, Nebot M et al. Bladder cancer and coffee consumption in smokers and non-smokers in Spain. Int J Epidemiol. 1993;22(1):38-44. doi:10.1093/ije/22.1.38.

64. Vena JE, Freudenheim J, Graham S, Marshall J, Zielezny M, Swanson M et al. Coffee, cigarette smoking, and bladder cancer in western New York. Annals of epidemiology. 1993;3(6):586-91. doi:10.1016/1047-2797(93)90079-j.

65. D'Avanzo B, La Vecchia C, Franceschi S, Negri E, Talamini R, Buttino I. Coffee consumption and bladder cancer risk. Eur J Cancer. 1992;28a(8-9):1480-4. doi:10.1016/0959-8049(92)90548-g.

66. Hashemian M, Sinha R, Murphy G, Weinstein SJ, Liao LM, Freedman ND et al. Coffee and tea drinking and risk of cancer of the urinary tract in male smokers. Annals of epidemiology. 2019;34:33-9. doi:10.1016/j.annepidem.2019.03.014.

67. Loftfield E, Freedman ND, Inoue-Choi M, Graubard BI, Sinha R. A Prospective Investigation of Coffee Drinking and Bladder Cancer Incidence in the United States. *Epidemiology* 2017;28(5):685-693. doi:10.1097/ede.0000000000000676.

68. Kuhlmann W, Fromme HG, Heege EM, Ostertag W. The mutagenic action of caffeine in higher organisms. Cancer Res. 1968;28(11):2375-89. doi:PMID: 4881507.

69. Zhang ZW, Xiao J, Luo W, Wang BH, Chen JM. Caffeine Suppresses Apoptosis of Bladder Cancer RT4 Cells in Response to Ionizing Radiation by Inhibiting Ataxia Telangiectasia Mutated-Chk2-p53 Axis. Chin Med J (Engl). 2015;128(21):2938-45. doi:10.4103/0366-6999.168065.

70. Sarkaria JN, Busby EC, Tibbetts RS, Roos P, Taya Y, Karnitz LM et al. Inhibition of ATM and ATR kinase activities by the radiosensitizing agent, caffeine. Cancer Res. 1999;59(17):4375-82.

71. Turati F, Bosetti C, Polesel J *et al*. Coffee, Tea, Cola, and Bladder Cancer Risk: Dose and Time Relationships. *Urology* 2015;86(6):1179-1184. doi:10.1016/j.urology.2015.09.017.

72. Sugiyama K, Sugawara Y, Tomata Y, Nishino Y, Fukao A, Tsuji I. The association between coffee consumption and bladder cancer incidence in a pooled analysis of the Miyagi Cohort Study and Ohsaki Cohort Study. *Eur J Cancer Prev* 2017;26(2):125-130. doi:10.1097/cej.0000000000000226.

73. Slattery ML, West DW, Robison LM. Fluid intake and bladder cancer in Utah. International journal of cancer. 1988;42(1):17-22. doi:10.1002/ijc.2910420105.

74. Cavin C, Holzhaeuser D, Scharf G, Constable A, Huber WW, Schilter B. Cafestol and kahweol, two coffee specific diterpenes with anticarcinogenic activity. Food Chem Toxicol. 2002;40(8):1155-63. doi:10.1016/s0278-6915(02)00029-7.

75. Huber WW, Scharf G, Nagel G, Prustomersky S, Schulte-Hermann R, Kaina B. Coffee and its chemopreventive components Kahweol and Cafestol increase the activity of O6-methylguanine-DNA methyltransferase in rat liver--comparison with phase II xenobiotic metabolism. Mutat Res. 2003;522(1-2):57-68. doi:10.1016/s0027-5107(02)00264-6.

76. Bravi F, Bosetti C, Tavani A, Bagnardi V, Gallus S, Negri E et al. Coffee drinking and hepatocellular carcinoma risk: a meta-analysis. Hepatology. 2007;46(2):430-5. doi:10.1002/hep.21708.

77. Bravi F, Scotti L, Bosetti C, Gallus S, Negri E, La Vecchia C et al. Coffee drinking and endometrial cancer risk: a metaanalysis of observational studies. Am J Obstet Gynecol. 2009;200(2):130-5. doi:10.1016/j.ajog.2008.10.032.

78. Je Y, Liu W, Giovannucci E. Coffee consumption and risk of colorectal cancer: a systematic review and meta-analysis of prospective cohort studies. International journal of cancer. 2009;124(7):1662-8. doi:10.1002/ijc.24124.

79. Turati F, Galeone C, La Vecchia C, Garavello W, Tavani A. Coffee and cancers of the upper digestive and respiratory tracts: meta-analyses of observational studies. Annals of oncology : official journal of the European Society for Medical Oncology. 2011;22(3):536-44. doi:10.1093/annonc/mdq603.

80. Ames BN, Gold LS. The causes and prevention of cancer: gaining perspective. Environmental health perspectives. 1997;105 Suppl 4:865-73. doi:10.1289/ehp.97105s4865.

81. Asaad NA, Zeng ZC, Guan J, Thacker J, Iliakis G. Homologous recombination as a potential target for caffeine radiosensitization in mammalian cells: reduced caffeine radiosensitization in XRCC2 and XRCC3 mutants. Oncogene. 2000;19(50):5788-800. doi:10.1038/sj.onc.1203953.

82. Saiki S, Sasazawa Y, Imamichi Y, Kawajiri S, Fujimaki T, Tanida I et al. Caffeine induces apoptosis by enhancement of autophagy via PI3K/Akt/mTOR/p70S6K inhibition. Autophagy. 2011;7(2):176-87. doi:10.4161/auto.7.2.14074.

83. Cortez D. Caffeine inhibits checkpoint responses without inhibiting the ataxia-telangiectasia-mutated (ATM) and ATM- and Rad3-related (ATR) protein kinases. J Biol Chem. 2003;278(39):37139-45. doi:10.1074/jbc.M307088200.

84. Zhou Y, Tian C, Jia C. A dose-response meta-analysis of coffee consumption and bladder cancer. *Prev Med* 2012;55(1):14-22. doi:10.1016/j.ypmed.2012.04.020.

85. Wu W, Tong Y, Zhao Q, Yu G, Wei X, Lu Q. Coffee consumption and bladder cancer: a meta-analysis of observational studies. *Sci Rep* 2015;5:1-9. doi:10.1038/srep09051.

86. Bracken MB, Triche E, Grosso L, Hellenbrand K, Belanger K, Leaderer BP. Heterogeneity in assessing self-reports of caffeine exposure: implications for studies of health effects. Epidemiology (Cambridge, Mass). 2002;13(2):165-71. doi:10.1097/00001648-200203000-00011.

87. Bracken MB, Triche E, Grosso L, Hellenbrand K, Belanger K, Leaderer BP. Heterogeneity in assessing self-reports of caffeine exposure: implications for studies of health effects. *Epidemiology* 2002;13(2):165-171.

88. Barry D. Differential recall bias and spurious associations in case/control studies. Stat Med. 1996;15(23):2603-16. doi:10.1002/(sici)1097-0258(19961215)15:23<2603::aid-sim371>3.0.co;2-g.

89. Barry D, Livingstone V. The investigation and correction of recall bias for an ordinal response in a case-control study. Stat Med. 2006;25(6):965-75. doi:10.1002/sim.2238.

90. Chouinard E, Walter S. Recall bias in case-control studies: an empirical analysis and theoretical framework. Journal of clinical epidemiology. 1995;48(2):245-54. doi:10.1016/0895-4356(94)00132-a.

91. Gefeller O. Invited commentary: Recall bias in melanoma -- much ado about almost nothing? Am J Epidemiol. 2009;169(3):267-70; discussion 71-2. doi:10.1093/aje/kwn362.

Supplementary Materials to Chapter 3

Supplementary Table 3-1 Characteristics of the Study Population (1 235 Cases and 1 783 Controls) and Caffeinated or Decaffeinated Coffee Consumption According to Age, Sex and Smoking Status

Supplementary Table 3-2 Ever versus Never Analyses for Caffeinated and Decaffeinated Coffee with Bladder Cancer Risk

Supplementary Table 3-3 Adjusted Odds Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption in Never Smokers

Supplementary Figure 3-1 Forest Plot of Meta-Analyses with ORs and 95% CIs for Ever and Never Coffee Consumption with Bladder Cancer Risk Adjusted for Age, Sex and Smoking

Supplementary Figure 3-2 Dose-response Relationships between Coffee Consumption and the Risk of Bladder Cancer among Never Smokers: a) Population-based Case-control Studies; b) Hospital-based Case-controls Studies

Supplementary Table 3-1 *Characteristics of the Study Population (1,235 Cases and 1,783 Controls) and Caffeinated or Decaffeinated Coffee Consumption According to Age, Sex and Smoking Status*

Characteristics		Coffee Consumption			Mean (±SD) cups/day	P [a]
		Never	Decaffeinated coffee	Caffeinated coffee		
N (%)		1,857 (73.23)	247 (9.74)	432 (17.03)	3.69 (±4.39)	0.001
Case (%)		751 (71.66)	129 (12.31)	168 (16.03)	4.30 (±3.59)	
Control (%)		1,106 (74.33)	118 (7.93)	264 (17.74)	3.28 (±4.40)	
Sex						
Male (%)	Case	632 (72.56)	92 (7.55)	147 (16.88)	4.20 (±3.65)	0.02
	Control	814 (72.29)	85 (7.55)	227 (20.16)	3.59 (±4.69)	
Female (%)	Case	119 (67.23)	37 (20.90)	21 (11.86)	4.79 (±3.26)	<0.001
	Control	292 (80.66)	33 (9.12)	37 (10.22)	2.33 (±4.34)	
Age (Mean ± SD) [b]	Case	63.16 (±8.39)	65.84 (±8.27)	65.43 (±11.34)	4.30 (±3.59)	<0.001
	Control	59.38 (±10.39)	65.18 (±8.48)	66.64 (±9.86)	3.28 (±4.40)	
Smoking Status						
Current (%)	Case	356 (77.39)	42 (9.13)	62 (13.48)	4.02 (±3.08)	0.01
	Control	356 (84.16)	18 (4.26)	49 (11.58)	2.19 (±4.96)	
Former (%)	Case	266 (64.25)	63 (15.22)	85 (20.53)	4.99 (±3.76)	0.04
	Control	333 (64.04)	55 (10.58)	132 (25.38)	4.55 (±4.30)	
Never (%)	Case	129 (74.14)	24 (13.79)	21 (12.07)	3.38 (±3.72)	0.07

[a] Calculated by χ^2 -test for categorical variables and t-test for continuous variables.

[b] Age at the time of recruitment.

Abbreviation: N, Number; SD, Standard Deviation.

P [a]: P-value <0.05 was considered statistically.

Supplementary Table 3-2 *Analyses for Caffeinated and Decaffeinated Coffee with Bladder Cancer Risk Adjusted for Age, Sex and Smoking*

Study Design	Coffee Consumption (ORs and 95% CI)		
	Never	Decaffeinated [c, d]	Caffeinated [c, d]
Caffeinated versus Decaffeinated		Ref.	1.88 (1.42-2.48)
Caffeinated versus Never	Ref.		1.52 (1.06-2.21)
Decaffeinated versus Never	Ref.	0.76 (0.29-2.03)	

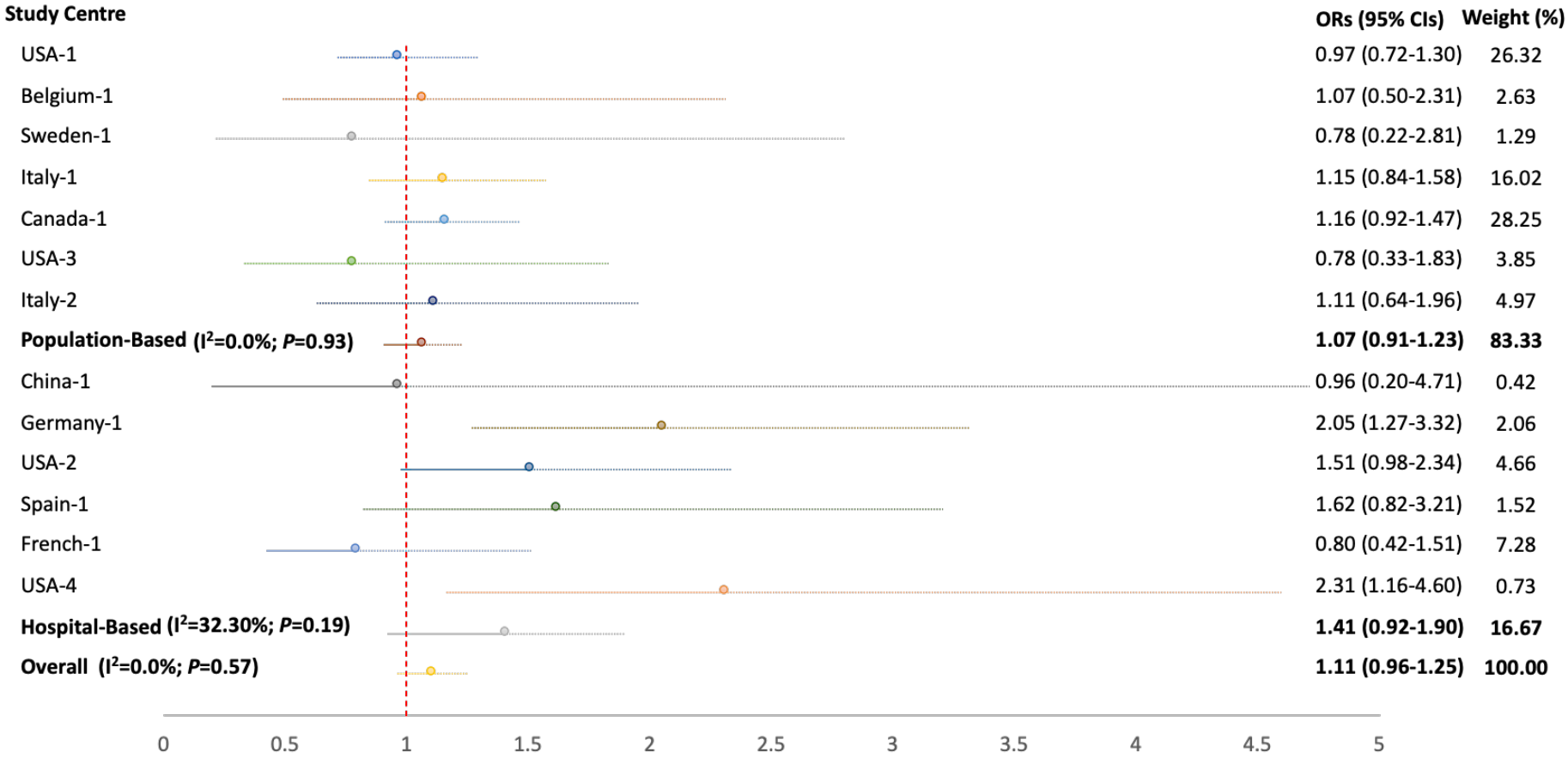
[c, d] *Model 2: Adjusted for age (years, continuous), sex (male or female) and smoking [if applicable, Smoking was defined as: 0 (never smokers); 1 (current light smokers (i.e. smoking less than 20 pack-years)); 2 (current heavy smokers (i.e. smoking more than 20 pack-years)); 3 (former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)); 4 (former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)].*
Abbreviation: OR, odds ratio; CI, confidence interval.
Reference group was never coffee consumers or decaffeinated coffee consumers.

Supplementary Table 3-3 *Adjusted Odds Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption in Never Smokers*

Models		Coffee Consumption (ORs and 95% CI)						P-trend
		Never	≤1 cup/day	1-2 cups/day	2-3 cups/day	3-4 cups/day	> 4 cups/day	
Unified Cup Size (237ml /cup)	Model 1 [c]	Ref.	0.97 (0.80-1.18)	1.40 (1.15-1.70)	1.14 (0.95-1.38)	1.50 (1.24-1.81)	1.76 (1.51-2.06)	0.05
	Model 2 [c, d]	Ref.	0.83 (0.68-1.02)	1.16 (0.95-1.43)	0.92 (0.76-1.12)	1.14 (0.94-1.40)	1.25 (1.06-1.47)	0.18
	Model 3 [c, d, e]	Ref.	0.85 (0.69-1.04)	1.12 (0.90-1.39)	0.98 (0.80-1.20)	1.30 (1.06-1.60)	1.56 (1.32-1.85)	0.11

[c] *Model 1: Crude model without adjustment.*
[c, d] *Model 2: Adjusted for age (years, continuous), sex (male or female) and smoking [if applicable, Smoking was defined as: 0 (never smokers); 1 (current light smokers (i.e. smoking less than 20 pack-years)); 2 (current heavy smokers (i.e. smoking more than 20 pack-years)); 3 (former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)); 4 (former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)].*

[c, d, e] Model 3: Additionally, adjusted for water consumption [millilitre (ml)/day], liquid milk consumption (ml/day), alcohol consumption (ml/day), carbonated drinks consumption (ml/day), tea consumption (ml/day) and juice consumption (ml/day).
Abbreviation: OR, odds ratio; CI, confidence interval; ml, millilitre.
Reference group was never coffee consumers.
P-trend <0.05 was considered statistically significant.



Supplementary Figure 3-1 *Forest Plot of Meta-Analyses with ORs and 95% CIs for Ever and Never Coffee Consumption with Bladder Cancer Risk Adjusted for Age, Sex and Smoking by Type of Control Selection*

Abbreviation: OR, odds ratio; CI, confidence interval.

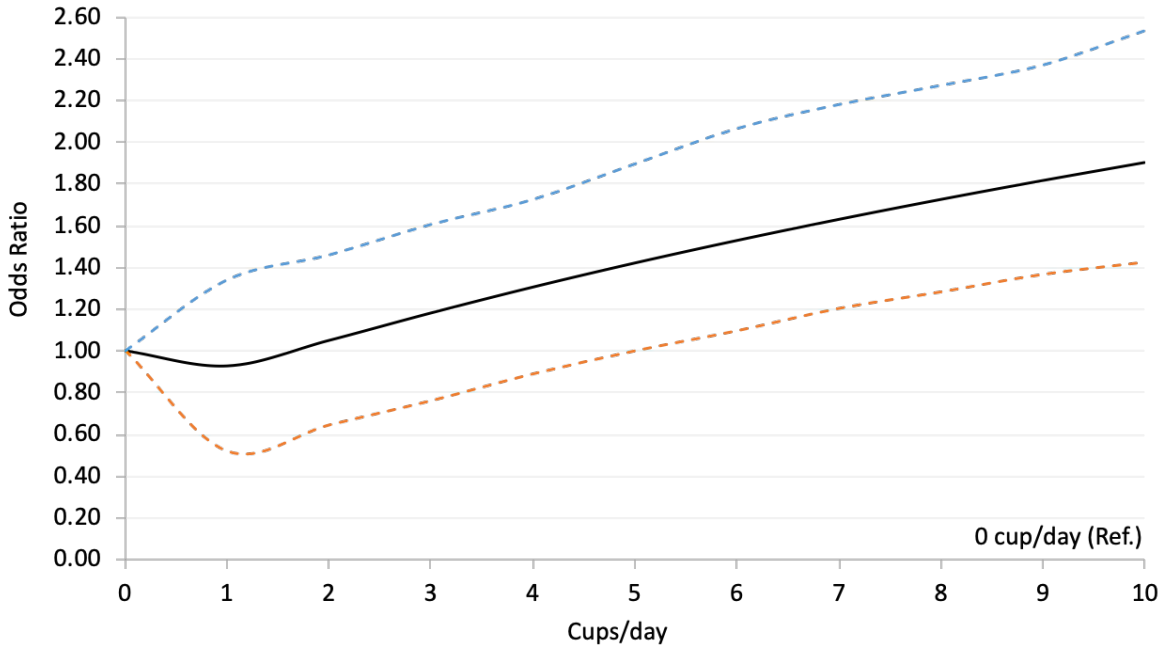
Circle dots denote the odds ratios (ORs); Horizontal lines represent the 95% confidence intervals (CIs); Weights are from random effects analyses.

Population-Based*: pooled OR of population-based case-control studies;* ***Hospital-Based****: pooled OR of hospital-based case-control studies;* ***Overall****: pooled OR of all studies.*

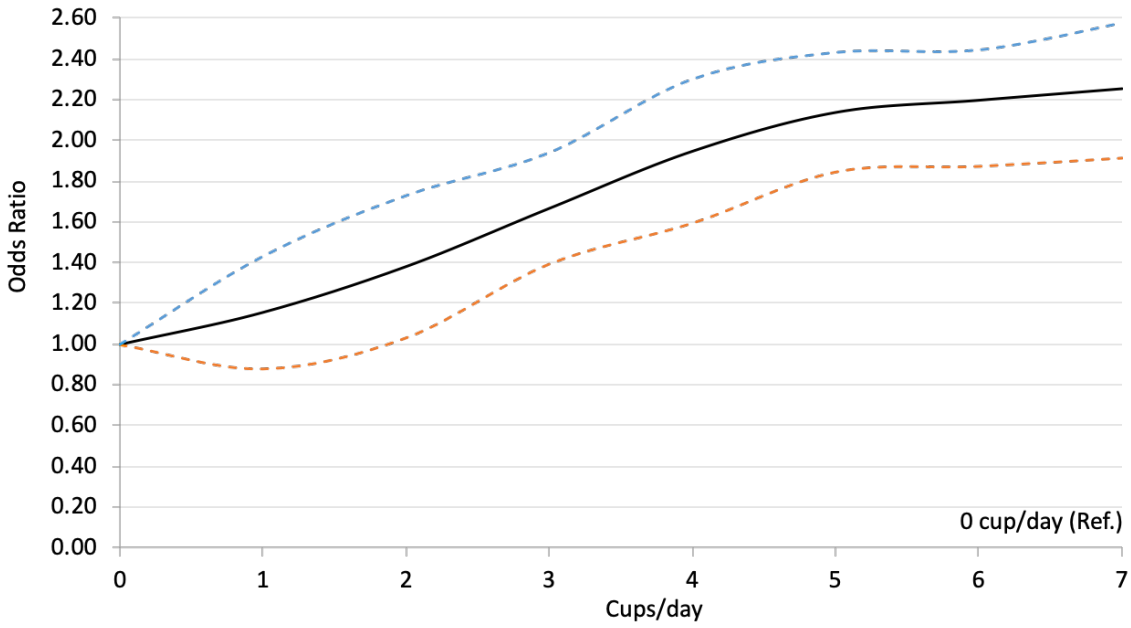
Reference group was non coffee consumers.

P-value <0.05 was considered statistically heterogeneous.

a) Population-based Case-control Studies
Non-linear Test: *P*-test=0.87
1 Cup/day Increment Risk: $OR_{-increase}$=1.06, 95% CI=1.03-1.08
P-increase<0.001



b) Hospital-based Case-control Studies
Non-linear Test: *P*-test=0.25
1 Cup/day Increment Risk: $OR_{-increase}$=1.13, 95% CI=1.10-1.15
P-increase<0.001



Supplementary Figure 3-2 *Dose-response Relationships between Coffee Consumption and the Risk of Bladder Cancer among Never Smokers: a) Population-based Case-control Studies; b) Hospital-based Case-control Studies*

The solid lines represent the odds ratios (ORs). The dashed lines represent the 95% confidence intervals (CIs) for the trend. The ORs were adjusted for age, sex and smoking (in the overall study population) (model 2).

Abbreviation: OR, odds ratio; CI, confidence interval.

P-test <0.05 was considered statistically significant.

P-increase <0.05 was considered statistically significant.

Chapter 4

Coffee Consumption and Bladder Cancer Risk: A Pooled Analysis of 501,604 Participants from 12 Prospective Cohort Studies

Evan Y.W. Yu, Yanan Dai, Anke Wesselius, Frits van Osch, Maree Brinkman, Piet van den Brandt, Eric J. Grant, Emily White, Elisabete Weiderpass, Marc Gunter, Bertrand Hemon, Maurice P. Zeegers

Adapted from European Journal of Epidemiology 2020 January

Abstract

Background: Recent epidemiological studies have shown varying associations between coffee consumption and bladder cancer. This research aims to elucidate the association between coffee consumption and bladder cancer risk by bringing together worldwide cohort studies on this topic.

Methods: Coffee consumption in relation to bladder cancer risk was examined by pooling individual data from 12 cohort studies, comprising of 2,601 cases out of 501,604 participants. Pooled multivariate *hazard ratios* (HRs), with corresponding 95% *confidence intervals* (CIs), were obtained using multilevel Weibull regression models. Furthermore, dose-response relationships were examined using *generalized least squares* (GLS) regression models.

Results: The association between coffee consumption and bladder cancer risk showed interaction with sex (*P*-interaction <0.001) and smoking (*P*-interaction=0.001). Therefore, analyses were stratified by sex and smoking. After adjustment for potential confounders, an increased bladder cancer risk was shown for high (>500 ml/day, equivalent to >4 cups/day) coffee consumption compared to never consumers among male smokers (current smokers: HR=1.75, 95% CI=1.27-2.42, *P*-trend=0.002; former smokers: HR=1.44, 95% CI=1.12-1.85, *P*-trend=0.001). In addition, dose-response analyses, in male smokers also showed an increased bladder cancer risk for coffee consumption of more than 500 ml/day (4 cups/day), with the risk of one cup (125 ml) increment as 1.07 (95% CI 1.06-1.08).

Conclusion: This research suggests that positive associations between coffee consumption and bladder cancer among male smokers but not never smokers and females. The inconsistent results between sexes and the absence of an association in never smokers indicate that the associations found among male smokers is unlikely to be causal and is possibly caused by residual confounding of smoking.

Key Words: Bladder Cancer; Coffee Consumption; Smoking; Dose-response Analyses; Cohort Study

Introduction

Bladder cancer is the 10th most common form of cancer worldwide, with an estimated 549,000 new cases and 200,000 deaths according to the latest global cancer statistics (2018) [1, 2]. Incidence rates of bladder cancer are highest in Europe and North America, with a strong predominance in males and the elderly [2-7]. Bladder cancer is reported to be the most expensive of all cancer in terms of life-time treatment due to its high rate of recurrence [8]. The strongest risk factors for bladder cancer occurrence, such as certain occupational exposures to chemical, water contaminants, and smoking [9, 10] have long been identified. However, as the bladder is an excretory organ, the role of fluid consumption could also be important in the development of bladder cancer.

Coffee is one of the most consumed beverages around the world. Since early 1970s [11], many studies have been conducted to detect the association between coffee consumption and bladder cancer risk; however, inconsistent findings have ranged from null associations [12-16], inverse associations [17, 18] to positive associations [19-25]. A Monographs Programme by *International Agency for Research on Cancer* (IARC) in 2016 reviewed the accumulated evidence of the coffee carcinogenicity focusing on bladder cancer risk and concluded: “no consistent evidence of an association between drinking coffee and bladder cancer risk” [26]. Therefore, they highlighted the need for large prospective studies with sufficient case numbers and detailed information on coffee consumption to better understand the association of this beverage and bladder cancer risk, especially properly controlling for smoking which is shown to be strongly associated with coffee consumption [27-32]. Furthermore, it is unclear and worthwhile to detect the threshold of coffee consumption above which a negative effect on bladder cancer risk might occur.

Our previous study, based on 13 case-control studies with 5,900 cases, yielded positive associations between heavy coffee consumption and bladder cancer among never smokers but not smokers [33]. Although this previous study had enough statistical power to provide definitive answers, it is commonly believed that results from case-control studies are unreliable due to a frequently occurring form of bias, namely; recall bias and selection bias. The present study, therefore, aims to replicate our previous findings based on BLEND data

originated from 12 cohort studies from around the world and give definitive answers on the influence of coffee consumption on bladder cancer risk.

Methods

Study Population

Data were derived from the *BLadder cancer Epidemiology and Nutritional Determinants* study (BLEND). BLEND is a large international epidemiology consortium that currently consists of 19 case-control studies and 16 cohort studies, which aims to pool data from available epidemiological studies on diet and bladder cancer. For the present study, 12 cohort studies (including 2,601 cases/499,003 non-cases), originated from 11 different countries in 3 continents [*i.e.* Europe: *European Prospective Investigation into Cancer and Nutrition* cohort studies (EPIC) [34] (France [35], Germany [36], Greece [37], Italy [38], the Netherlands [39], Norway [40], Spain [37], Sweden [41, 42], United Kingdom [43, 44]], *NetherLands Cohort Study* (NLCS) [45]; North America: *VITamins And Lifestyle* cohort study (VITAL) [46]; and Asia: *Radiation Effects Research Foundation-Life Span Study* (RERF-LSS) [47]], had sufficient information on coffee consumption to be eligible for inclusion and each study was ethically approved [34, 45, 46, 48] (**Supplementary Table 4-1**).

Data Collection and Coding

Details on the methodology of the BLEND consortium have been described elsewhere [49]. All included studies used a self-administered or trained interviewer administered *food frequency questionnaire* (FFQ) that was validated on either food groups [34, 46, 50-52], and/or energy intake [51, 53]. For each study, participants were asked to report on their usual coffee consumption during the preceding year before study enrolment. The collected coffee consumption were harmonized and categorized by using the hierarchal Eurocode 2 food coding system developed by the European Union [54], besides, weekly, monthly or yearly coffee consumption were converted to daily *millilitres* (ml) of coffee consumption. Person-years of follow-up for each participant was calculated from date of study enrolment until date of bladder cancer diagnosis, or date of ending follow up (*e.g.* date of death, lost to follow-up, or study exit), whichever came first. For the NLCS study, a nested case-cohort approach was applied, in which the number of person-years at risk was estimated based on a sub-cohort that was randomly sampled [45].

Each study ascertained incident bladder cancer, defined to include all urinary bladder neoplasms according to the *International Classification of Diseases for Oncology* (ICD-O-3 code C67) using population-based cancer registries, health insurance records, or medical records. Bladder cancers were classified as non-muscle invasive bladder cancer (NMIBC) or muscle invasive bladder cancer (MIBC). NMIBC included non-invasive papillary carcinomas confined to the urothelium (stage Ta), and carcinomas that invaded the lamina propria of the bladder wall (stage T1). High grade flat non-invasive carcinomas confined to the urothelium (*carcinoma in situ*; CIS) without other concomitant tumour stages [*i.e.* T1/Ta (classified to non-muscle invasive prior), or muscle invasive] were also classified as NMIBC. MIBC included carcinomas that invaded into the detrusor muscle (stage T2), carcinomas that invaded into the peri-vesical tissue (stage T3), and carcinomas that invaded adjacent tissues and organs (most often the prostate or uterus, stage T4).

In addition to information on coffee and other dietary intake data, the BLEND dataset also included data on study characteristics (design, method of dietary assessment, geographical region), participant demographics (age, sex and ethnicity), smoking status and smoking pack-years (*i.e.* the number of cigarettes smoked per day multiplying the years of smoking), which were measured at baseline.

Statistical Analyses

The differences between cases and non-cases were examined by Chi-square test for categorical variables and t-test for continuous variables. To assess the influence of coffee consumption on the bladder cancer risk, multilevel (2-level) Weibull regression analyses were used to estimate the pooled *hazard ratios* (HRs) and 95% *confidence intervals* (CIs), which nested the individuals within study centres to adjust for cross-cohort heterogeneity. We found no violation for the appropriateness of the lognormal and Weibull model by using Wald test. Coffee consumption was divided into a never consumption group and 3 ever consumption groups based on a tertile ordered distribution: low consumption (0-180 ml/day), medium consumption (180-500 ml/day) and high consumption (>500 ml/day). The Weibull regression models used never coffee consumers as the reference group and associations were computed as model 1: adjusted for age (years) and sex (male or female), model 2: adjusted

for mode 1 and smoking (status-pack-years, smoking was defined as: 0 (never smokers); 1 [current light smokers (*i.e.* smoking less than 20 pack-years)]; 2 [current heavy smokers (*i.e.* smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (*i.e.* smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (*i.e.* smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), or model 3: adjusted for model 2 and total energy intake [*kilocalorie* (kcal)/day, continuous]), and other potential confounders that showed to be different between bladder cancer cases and non-bladder cancer cases; which included tea consumption (ml/day, continuous; another popular and main caffeine-source beverage), other caffeine intake [*gram* (g)/day, continuous; residual caffeine content from other foods and beverages], and sweeteners intake (g/day, continuous; popular additive accompanied with coffee consumption). The effects of caffeinated or decaffeinated coffee on bladder cancer risk were assessed by comparing to never coffee consumers based on multilevel Weibull regression (model 3). To understand the relevance of the effect modification, the main interaction terms between coffee consumption and age, sex and smoking were added to the model 2. *P*-interaction <0.05 was considered statistically significant where upon all analyses were stratified for the covariate of interest. Stratification analyses were assessed regarding bladder cancer subtypes (*i.e.* NMIBC and MIBC), sex and smoking. Several sensitivity analyses were performed: a) participants who are neither smokers nor coffee consumers were used as a reference group to investigate the bladder cancer risk with coffee consumption in different smoking status; b) bladder cancer cases diagnosed within the first 2 years after recruitment were excluded. Missing variables (*e.g.* total energy intake and tea consumption) were imputed separately in each participating cohort by multiple imputation method considering different study protocols and characteristics of participants across cohorts. Only participants with complete information on bladder cancer status (*i.e.* cases or non-cases), age, sex and smoking were included in building of the imputation models. Linear regression models were fitted for those two continuous variables (total energy intake and tea consumption) respectively. After imputation, one single complete dataset was created and used to estimate results based on model 3.

In our secondary analysis, a potential dose-response relationship between coffee consumption and bladder cancer was assessed by using *generalized least squares* (GLS) regression models [55]. Restricted cubic splines, which set knots at the 5th, 35th, 65th and 95th percentile, were used to investigate statistical linearity or non-linearity for all curves [56]. The results of the dose-response analyses were presented for each regular European cup (125 ml) of coffee increment up to 10 cups of coffee per day (1,250 ml/day) compared to never coffee consumers. Adjustments were made for age, sex, smoking, total energy intake, tea consumption, other caffeine intake, and sweeteners intake. As a last step, the associations of (highest vs. never) coffee consumption with bladder cancer risk in overall population, male smokers, male never smokers, female smokers and female never smokers, were assessed is each participating cohort separately and combined in a meta-analysis approach using a random-effect model and adjusting for model 3.

All statistical analyses were performed with STATA version 14 SE (Stata Corporation, Texas, USA). Two-tailed probabilities (P values) <0.05 were considered statistically significant.

Results

Baseline Characteristics

The baseline characteristics of the study population are shown in **Table 4-1**. Altogether, 2,601 incident cases of bladder cancer (1,940 males, 661 females) and 499,003 non-cases were included in our analyses, with a total of 5,373,352 person-years of follow-up (median follow-up: 12 years). Of the cases, only 1,617 (62%) were recorded diagnoses by NMIBC (38%) and MIBC (24%). The median age at baseline was 61 years for cases and 52 years for non-cases. The majority (97.94%) of participants were Caucasian, with only 2.06% non-Caucasian (*i.e.* Asian, Indian and Black). Approximately 23% of participants reported drinking coffee more than 500 ml per day, with an average consumption of 789 ml/day. Among those heavy consumers, most of them (64%) reported a smoking history, with 35% being former smokers. At baseline, a higher coffee consumption was observed among smokers (446 ml/day) compared to never smokers (326 ml/day), as well as among males (430 ml/day) compared to females (369 ml/day).

Table 4-1 *Characteristics of the Study Population (2,601 cases and 499,003 non-cases) and Coffee Consumption Amount of Included Studies*

Characteristics		N (%)					Person-years
			Case (%)	NMIBC (%)	MIBC (%)	Non-case (%)	
	Total	501,604	2,601	994	623	499,003	5,373,352
	Never	126,875 (25.29)	405 (15.57)	140 (14.08)	39 (6.26)	126,470 (25.34)	1,376,399
Coffee Consumption	Low ≤180 ml/day	125,621 (25.04)	394 (15.15)	135 (13.58)	54 (8.67)	125,227 (25.10)	1,409,850
	Medium 180-500 ml/day	132,731 (26.46)	887 (34.10)	321 (32.29)	246 (39.49)	131,844 (26.42)	1,402,186
	High >500 ml/day	116,377 (23.20)	915 (35.18)	398 (40.04)	284 (45.58)	115,462 (23.14)	1,184,917
	Mean [ml/day (SD)]	388.41 (340.00)	505.55 (343.32)	527.20 (337.60)	586.89 (342.55)	387.72 (339.86)	388.41 (340.00)
P [a]		<0.001					<0.001
P [b]							

Table 4-1_continued *Characteristics of the Study Population (2,601 cases and 499,003 non-cases) and Coffee Consumption Amount of Included Studies*

Characteristics		**Sex**				**Age [Mean (SD)]**	
		Male (%)		Female (%)			
		Case	Non-case	Case	Non-case	Case	Non-case
	Total	1,940	153,204	661	345,799	2,601	499,003
	Never	283 (14.59)	42,484 (27.73)	121 (18.31)	83,986 (24.29)	58.35 (8.58)	51.71 (10.25)
	Low ≤180 ml/day	281 (14.48)	36,512 (23.83)	113 (17.10)	88,715 (25.66)	58.92 (9.37)	50.70 (11.17)
Coffee Consumption	Medium 180-500 ml/day	652 (33.61)	34,167 (22.30)	235 (35.55)	98,677 (28.54)	62.54 (6.78)	52.90 (10.84)
	High >500 ml/day	724 (37.32)	40,041 (26.14)	192 (29.04)	75,421 (21.81)	60.92 (7.28)	52.81 (10.15)
	Mean [ml/day (SD)]	521.19 (348.96)	430.51 (386.72)	457.45 (320.93)	369.62 (316.24)	505.55 (343.32)	387.72 (339.86)
P [a]		<0.001		<0.001		<0.001	
P [b]		<0.001				0.281	

Table 4-1_continued *Characteristics of the Study Population (2,601 cases and 499,003 non-cases) and Coffee Consumption Amount of Included Studies*

Characteristics		Smoking Status					
		Never (%)		Current (%)		Former (%)	
		Case	Non-case	Case	Non-case	Case	Non-case
	Total	526	255,742	981	99,029	1,094	144,232
	Never	122 (23.19)	75,338 (29.46)	141 (14.37)	23,443 (23.67)	142 (12.98)	27,689 (19.20)
Coffee Consumption	Low ≤180 ml/day	100 (19.01)	71,086 (27.80)	142 (14.48)	21,436 (21.65)	152 (13.89)	32,705 (22.67)
	Medium 180-500 ml/day	199 (37.83)	67,046 (26.22)	274 (27.93)	21,722 (21.93)	414 (37.84)	43,076 (29.87)
	High >500 ml/day	105 (19.96)	42,272 (16.53)	424 (43.22)	32,428 (32.75)	386 (35.28)	40,762 (28.26)
	Mean [ml/day (SD)]	382.47 (275.87)	325.88 (298.32)	572.54 (375.08)	482.68 (391.39)	498.67 (324.40)	421.85 (345.23)
P [a]		<0.001		<0.001		<0.001	
P [b]				0.001			

Table 4-1_continued *Characteristics of the Study Population (2,601 cases and 499,003 non-cases) and Coffee Consumption Amount of Included Studies*

Characteristics		Smoking Pack-years [Mean (SD)]	Ethnicity (%)		Total Energy [kcal/day (SD)]	Tea [ml/day (SD)]	Other Caffeine [g/day (SD)]	Sweeteners [g/day (SD)]
			Caucasian	Non-Caucasian				
	Total	84,990	491,013	10,311	501,604	501,604	501,604	501,604
Coffee Consumption	Never	21.40 (17.62)	125,735 (25.61)	1,065 (10.33)	2,087.70 (706.82)	161.17 (278.91)	107.40 (279.88)	44.81 (96.15)
	Low ≤180 ml/day	19.86 (17.13)	119,297 (24.30)	6,276 (60.87)	2,683.39 (866.49)	242.83 (343.69)	344.30 (104.07)	37.15 (37.12)
	Medium 180-500 ml/day	20.02 (15.81)	131,989 (26.88)	690 (6.69)	2,039.71 (616.90)	283.47 (328.60)	359.94 (122.84)	36.90 (41.14)
	High >500 ml/day	23.58 (15.87)	113,992 (23.22)	2,280 (22.11)	2,245.24 (1592.76)	214.56 (280.22)	551.60 (174.57)	33.26 (44.27)
	Mean [ml/day (SD)]	352.99 (372.15)	392.72 (339.49)	215.21 (308.98)	388.41 (340.00)	388.41 (340.00)	388.41 (340.00)	388.41 (340.00)
P [a]		<0.001	0.06	0.35	<0.001	<0.001	<0.001	0.02
P [b]		0.001						

[a] *Calculated by Chi-square (χ^2) test for categorical variables and t-test for continuous variables between bladder cancer cases and non-cases. Age at the time of recruitment.*

Abbreviation: N, Number; SD, Standard Deviation; ml, millilitre; kcal, kilocalorie; g, gram.

P [a]: $P\text{-}_{value}$ *<0.05 was considered statistically significant.*

P [b]: $P\text{-}_{interaction}$ *<0.05 was considered statistically significant.*

Association between Coffee Consumption and Bladder Cancer Risk

Amount of coffee consumption with Bladder Cancer risk

The results of multilevel Weibull regressions for the association between coffee consumption and bladder cancer risk are shown in **Table 4-2**. Overall, compared to never coffee consumers, an increased bladder cancer risk was observed for coffee consumption of more than 500 ml/day (HR_{model3}=1.56, 95% CI=1.38-1.77, P-trend <0.001) and of 180-500 ml/day (HR_{model3}=1.39, 95% CI=1.23-1.58, P-trend <0.001). In addition, both NMIBC and MIBC showed similar results; an increased bladder cancer risk was observed for coffee consumption of more than 500 ml/day (HR_{model3}=1.39, 95% CI=1.11-1.74, P-trend <0.001; HR_{model3}=2.05, 95% CI=1.41-2.96, P-trend=0.21, respectively).

Coffee consumption with Bladder Cancer risk stratified by sex and smoking

Since coffee consumption showed interaction with sex (P-interaction <0.001) and smoking (P-interaction=0.001), analyses were stratified by sex and smoking and compared to never coffee consumers (**Table 4-3**). For males, according to model 3, a significant increased bladder cancer risk was observed in high (>500 ml/day) coffee consumption for both current and former smokers (current smokers: HR_{model3} =1.75, 95% CI =1.27-2.42, P-trend=0.002; former smokers: HR_{model3}=1.44, 95% CI =1.12-1.85, P-trend=0.001), while no association was found among male never smokers. Similar increased bladder cancer risks were observed for heavy (more than 20 pack-years) male smokers and light (less than 20 pack-years) male smokers when comparing high (>500 ml/day) coffee consumers to never consumers (heavy smokers: HR_{model3}=1.72, 95% CI =1.13-2.62, P-trend=0.02; light smokers: HR_{model3}=1.93, 95% CI=1.07-3.47, P-trend=0.03). In addition, the moderate (180-500 ml/day) coffee consumption of male current smokers and male heavy smokers also showed increased risks of bladder cancer (HR_{model3}=1.69, 95% CI=1.23-2.34, P-trend =0.002; HR_{model3}=1.60, 95% CI=1.05-2.43, P-trend =0.02, respectively). For females, no significant association or increased bladder cancer risk could be observed for neither smokers nor never smokers.

Table 4-2 *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Amount Stratified by NMIBC and MIBC*

Subgroup		**Overall**			**NMIBC**	
Models		Model 1 [b]	Model 2 [b, c]	Model 3 [b, c, d]	Model 1 [b]	Model 2 [b, c]
No. Cases/Total Participants		2,601/501,604	2,601/501,604	2,601/501,604	994/385,563	994/385,563
Coffee Consumption	Never	Ref.	Ref.	Ref.	Ref.	Ref.
	≤180 ml/day	1.01 (0.88-1.16)	1.01 (0.88-1.17)	1.03 (0.89-1.19)	0.75 (0.59-0.97)	0.77 (0.60-1.02)
	180-500 ml/day	1.70 (1.51-1.92)	1.38 (1.22-1.56)	1.39 (1.23-1.58)	1.05 (0.82-1.35)	1.36 (1.08-1.70)
	>500 ml/day	2.09 (1.85-2.35)	1.55 (1.37-1.75)	1.56 (1.38-1.77)	1.30 (1.03-1.65)	1.38 (1.11-1.72)
P-trend		<0.001	0.01	<0.001	0.001	<0.001

Table 4-2_continued *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Amount Stratified by NMIBC and MIBC*

Subgroup		NMIBC	MIBC		
Models		Model 3 [b, c, d]	Model 1 [b]	Model 2 [b, c]	Model 3 [b, c, d]
No. Cases/Total Participants		994/385,563	623/385,192	623/385,192	623/385,192
Coffee Consumption	Never	Ref.	Ref.	Ref.	Ref.
	≤180 ml/day	0.75 (0.49-1.02)	1.15 (0.75-1.76)	1.37 (0.90-2.09)	1.47 (0.96-2.24)
	180-500 ml/day	1.34 (1.06-1.68)	1.26 (0.87-1.84)	2.28 (1.58-3.31)	2.44 (1.68-3.55)
	>500 ml/day	1.39 (1.11-1.74)	1.62 (1.12-2.34)	2.00 (1.39-2.88)	2.05 (1.41-2.96)
P-trend		<0.001	0.001	0.27	0.21

[b] *Model 1: Adjusted for age (years, continuous), sex (male or female).*

[b, c] *Model 2: Additionally, adjusted for smoking (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]).*

[b, c, d] *Model 3: Additionally, adjusted for total energy intake (kcal/day, continuous), tea consumption (ml/day, continuous), other caffeine intake (g/day, continuous), and sweeteners intake (g/day, continuous).*

Abbreviation: NMIBC, Non-Muscle Invasive Bladder Cancer; MIBC, Muscle Invasive Bladder Cancer; ml, millilitre; kcal, kilocalorie; g, gram.

Reference group was never coffee consumers.

P-trend <0.05 was considered statistically significant.

Table 4-3 *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Amount Stratified by NMIBC and MIBC*

Subgroup		Male					
		Never Smoker			Current Smoker		
Models		Model 1 [b]	Model 2 [b, c1]	Model 3 [b, c1, d]	Model 1 [b]	Model 2 [b, c1]	Model 3 [b, c1, d]
No. Cases/Total Participants		248/54,609	248/54,609	248/54,609	786/38,526	786/38,526	786/38,526
Coffee Consumption	Never	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	≤180 ml/day	0.73 (0.48-1.12)	0.77 (0.52-1.15)	0.75 (0.49-1.15)	1.18 (0.88-1.59)	1.09 (0.82-1.46)	1.24 (0.92-1.67)
	180-500 ml/day	1.06 (0.73-1.54)	1.00 (0.68-1.45)	1.05 (0.72-1.55)	1.66 (1.21-2.30)	1.59 (1.17-2.16)	1.69 (1.23-2.34)
	>500 ml/day	0.90 (0.60-1.34)	0.85 (0.57-1.28)	0.89 (0.60-1.35)	1.78 (1.30-2.44)	1.69 (1.25-2.29)	1.75 (1.27-2.42)
P-trend		0.94	0.70	0.91	0.001	0.001	0.002

Table 4-3_continued *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Amount Stratified by NMIBC and MIBC*

Subgroup		Male					
		Former Smoker			Heavy Smoker [e]		
Models		Model 1 [b]	Model 2 [b, c1]	Model 3 [b, c1, d]	Model 1 [b]	Model 2 [b, c2]	Model 3 [b, c2, d]
No. Cases/Total Participants		906/62,009	906/62,009	906/62,009	706/17,778	706/17,778	706/17,778
Coffee Consumption	Never	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	≤180 ml/day	1.12 (0.85-1.46)	1.11 (0.84-1.45)	1.08 (0.82-1.42)	1.13 (0.77-1.66)	1.12 (0.76-1.65)	1.13 (0.77-1.67)
	180-500 ml/day	2.39 (1.87-3.06)	2.38 (1.86-3.04)	1.29 (0.99-1.67)	1.58 (1.04-2.38)	1.55 (1.03-2.35)	1.60 (1.05-2.43)
	>500 ml/day	1.87 (1.46-2.39)	1.86 (1.45-2.38)	1.44 (1.12-1.85)	1.71 (1.13-2.58)	1.73 (1.15-2.62)	1.72 (1.13-2.62)
P-trend		<0.001	<0.001	0.001	0.01	0.02	0.02

Table 4-3_continued *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Amount Stratified by NMIBC and MIBC*

Subgroup		Male			Female		
		Light Smoker [e]			Never Smoker		
Models		Model 1 [b]	Model 2 [b, c2]	Model 3 [b, c2, d]	Model 1 [b]	Model 2 [b, c1]	Model 3 [b, c1, d]
No. Cases/Total Participants		241/11,774	241/11,774	241/11,774	278/201,659	278/201,659	278/201,659
Coffee Consumption	Never	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	≤180 ml/day	1.26 (0.72-2.21)	1.28 (0.73-2.23)	1.44 (0.82-2.53)	0.92 (0.62-1.36)	0.96 (0.66-1.41)	0.95 (0.64-1.40)
	180-500 ml/day	1.58 (0.90-2.77)	1.55 (0.89-2.70)	1.67 (0.94-2.98)	0.95 (0.64-1.41)	0.98 (0.66-1.44)	0.97 (0.66-1.44)
	>500 ml/day	1.85 (1.05-3.27)	1.88 (1.07-3.30)	1.93 (1.07-3.47)	0.90 (0.58-1.38)	0.92 (0.60-1.41)	0.93 (0.60-1.44)
P-trend		0.03	0.02	0.03	0.67	0.73	0.79

Table 4.3_continued *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Amount Stratified by NMIBC and MIBC*

Subgroup		Female					
		Current Smoker			Former Smoker		
Models		Model 1 [b]	Model 2 [b, c1]	Model 3 [b, c1, d]	Model 1 [b]	Model 2 [b, c1]	Model 3 [b, c1, d]
No. Cases/Total Participants		195/61,484	195/61,484	195/61,484	188/83,317	188/83,317	188/83,317
Coffee Consumption	Never	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	≤180 ml/day	0.79 (0.46-1.35)	0.76 (0.45-1.30)	0.81 (0.47-1.38)	0.55 (0.32-0.93)	0.57 (0.33-0.97)	0.56 (0.33-0.95)
	180-500 ml/day	0.75 (0.45-1.28)	0.73 (0.43-1.23)	0.76 (0.45-1.29)	0.71 (0.46-1.10)	0.71 (0.45-1.14)	0.71 (0.46-1.10)
	>500 ml/day	0.95 (0.58-1.57)	1.03 (0.62-1.69)	0.94 (0.56-1.56)	0.74 (0.47-1.16)	0.73 (0.45-1.17)	0.74 (0.47-1.16)
P-trend		0.85	0.56	0.94	0.50	0.51	0.47

Table 4-3_continued *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Amount Stratified by NMIBC and MIBC*

Subgroup		Female					
		Heavy Smoker [e]			Light Smoker [e]		
Models		Model 1 [b]	Model 2 [b, c2]	Model 3 [b, c2, d]	Model 1 [b]	Model 2 [b, c2]	Model 3 [b, c2, d]
No. Cases/Total Participants		116/20,129	116/20,129	116/20,129	89/35,309	89/35,309	89/35,309
Coffee Consumption	Never	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	≤180 ml/day	0.88 (0.42-1.87)	0.88 (0.42-1.87)	0.89 (0.42-1.90)	0.60 (0.27-1.33)	0.59 (0.27-1.31)	0.60 (0.27-1.33)
	180-500 ml/day	0.93 (0.46-1.88)	0.93 (0.46-1.88)	0.90 (0.44-1.83)	0.51 (0.24-1.09)	0.50 (0.23-1.08)	0.48 (0.22-1.04)
	>500 ml/day	0.92 (0.46-1.85)	0.93 (0.46-1.85)	0.86 (0.42-1.76)	0.80 (0.38-1.71)	0.75 (0.35-1.62)	0.73 (0.33-1.59)
P-trend		0.87	0.88	0.68	0.91	0.81	0.73

[b] Model 1: Adjusted for age (years, continuous).

[b, c1] Model 2: Additionally, adjusted for smoking pack-years if applicable; [b, c2] Model 2: Additionally, adjusted for smoking status if applicable (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]).

[b, c1/c2, d] Model 3: Additionally, adjusted for total energy intake (kcal/day, continuous), tea consumption (ml/day, continuous), other caffeine intake (g/day, continuous), and sweeteners intake (g/day, continuous).

[e] Heavy Smoker: Pack-years >20; Light Smoker: Pack-years ≤20; and both heavy and light smokers included current and former smokers.

Abbreviation: ml, millilitre; kcal, kilocalorie; g, gram.

Reference group was never coffee consumers.

P-trend <0.05 was considered statistically significant.

Caffeinated and decaffeinated coffee with Bladder Cancer risk

Results for the comparison of caffeinated and decaffeinated coffee are shown in **Table 4-4**. Nine cohort studies [35-44], with a total of 310,116 participants (1,040 bladder cancer cases, 182,716 caffeinated coffee consumers, 18,058 decaffeinate coffee consumers, and 109,342 neither caffeinated or decaffeinated coffee consumers), were included in the analyses. For the analysis on caffeinated coffee, the consumption over 500 ml/day showed an increased bladder cancer risk compared to never coffee consumers based on model 3 (HR_{model3}=1.12, 95% CI=1.01-1.23, P-trend=0.46); here the significant association maintained in smokers (HR_{model3}=1.22, 95% CI=1.06-1.42, P-trend=0.10) while not in never smokers. Consumers of decaffeinated coffee suggested similar bladder cancer risks compared to consumers of caffeinated coffee. However, since the sample size for decaffeinated coffee decreased significantly, results should be interpreted with caution.

Dose-response Analyses

Dose-response relationships between coffee consumption and the risk of bladder cancer are displayed in **Figure 4-1**. The curves for the overall population showed slightly increased bladder cancer risks with significant increment for consuming over 500 ml/day (4 cups/day), even though the tendency went down marginally after 1,000 ml/day (8 cups/day). Similar results were found among male smokers with significant increased risks for consuming over 400 ml/day (3.2 cups/day). No significant dose-response relationships were observed for never smokers (**Figure 4-1**) and females (**Supplementary Figure 4-1**). Adjusted HRs (model 3) and 95% CIs for 125 ml/day (1 cup/day) increment were 1.04 (95% CI 1.03-1.05) in overall study population, and 1.07 (95% CI 1.06-1.08) in male smokers.

Table 4-4 *Hazard Ratios and 95% Confidence Intervals of Caffeinated and Decaffeinated Coffee Consumption with Bladder Cancer Risk Based on Model 3* [b, c, d]

Study Subgroup		Caffeinated Coffee			Decaffeinated Coffee		
		Overall population	Never smokers	Ever smokers [f]	Overall population	Never smokers	Ever smokers [f]
No. Cases/Total Participants		968/292,058	221/150,671	747/141,387	389/127,400	109/75,781	280/51,619
Coffee Consumption	Never	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	Ever	1.07 (0.92-1.23)	0.78 (0.57-1.07)	1.15 (0.95-1.40)	1.16 (0.87-1.55)	0.90 (0.49-1.66)	1.30 (0.93-1.81)
	≤180 ml/day	1.05 (0.88-1.25)	0.89 (0.60-1.32)	1.10 (0.87-1.40)	1.05 (0.75-1.47)	0.81 (0.39-1.67)	1.17 (0.80-1.71)
	180-500 ml/day	1.05 (0.88-1.25)	0.69 (0.46-1.03)	1.17 (0.93-1.47)	1.50 (0.90-2.50)	0.87 (0.12-1.50)	1.43 (0.55-2.73)
	>500 ml/day	1.12 (1.01-1.23)	0.79 (0.48-1.29)	1.22 (1.06-1.42)	1.15 (0.49-2.73)	1.13 (0.45-2.90)	1.78 (0.62-3.25)
P-trend		0.46	0.13	0.10	0.33	0.07	0.08

[b, c, d] *Model 3: Adjusted for age (years, continuous), sex (male or female), smoking (if applicable, smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), total energy intake (kcal/day, continuous), tea consumption (ml/day, continuous), other caffeine intake (g/day, continuous), and sweeteners intake (g/day, continuous).*

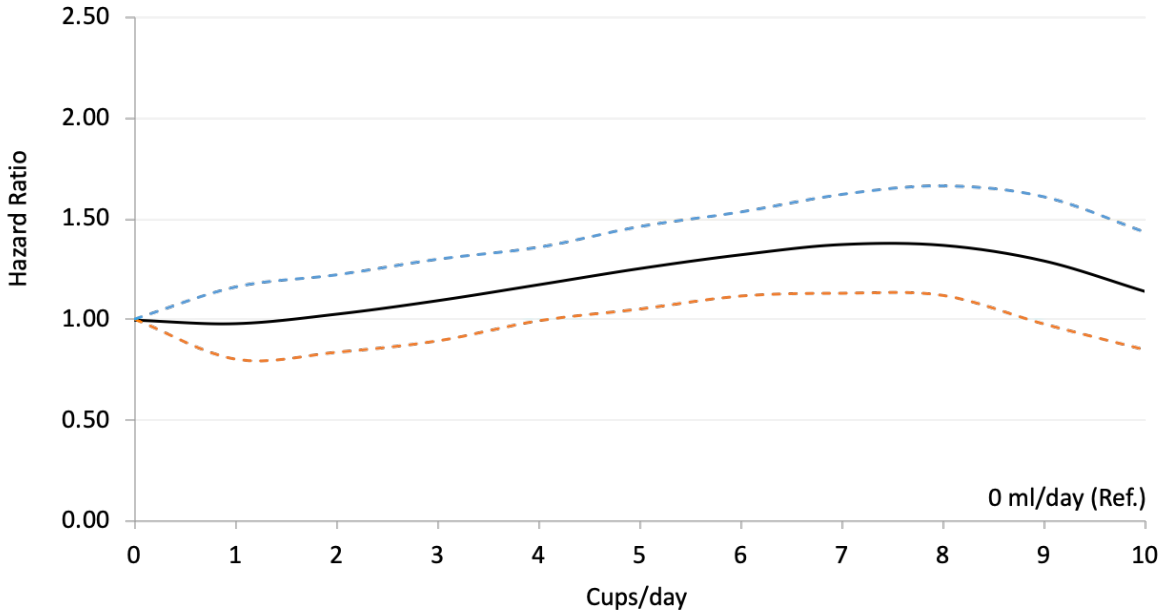
[f] *Ever smokers included both current/former smokers and heavy/light smokers.*

Abbreviation: ml, millilitre; kcal, kilocalorie; g, gram.

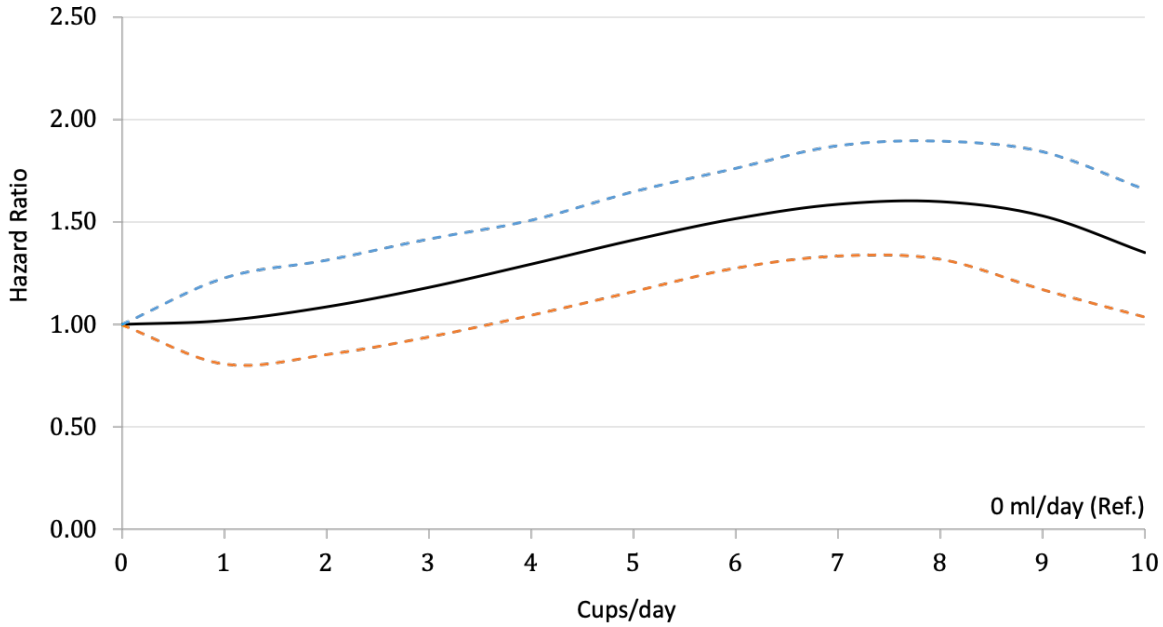
Reference group was never coffee consumers.

P-trend <0.05 was considered statistically significant.

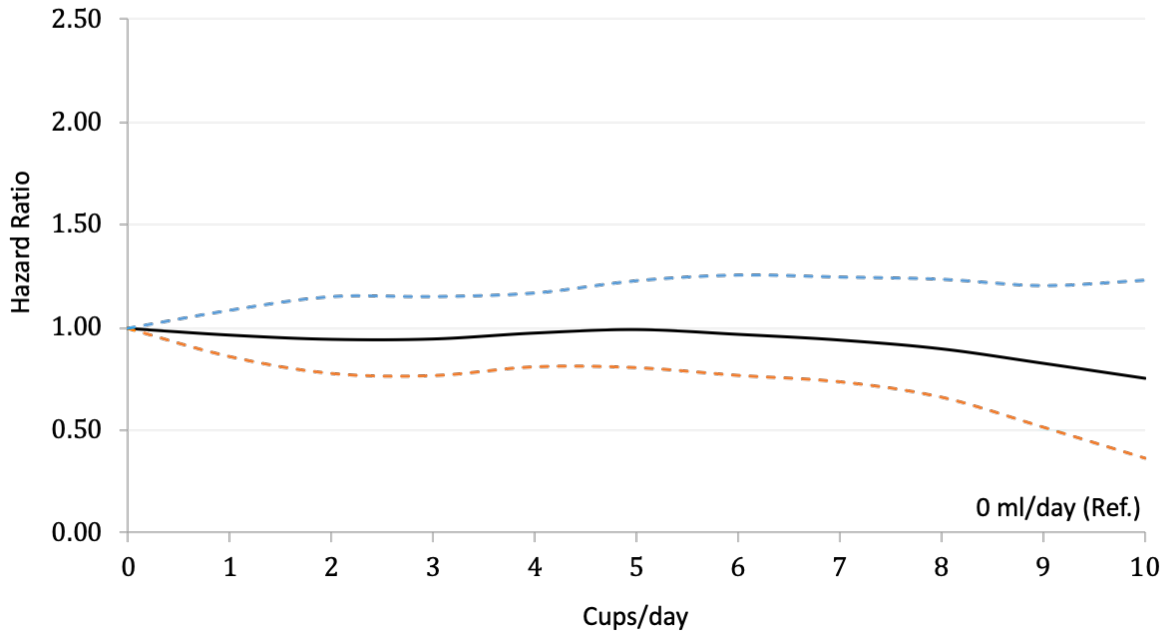
a) Overall Study Population
Non-linear Test: *P*-test=0.06
1 Cup/day Increment Risk: $HR_{-increase}$=1.04, 95% CI=1.03-1.05
P-increase<0.001



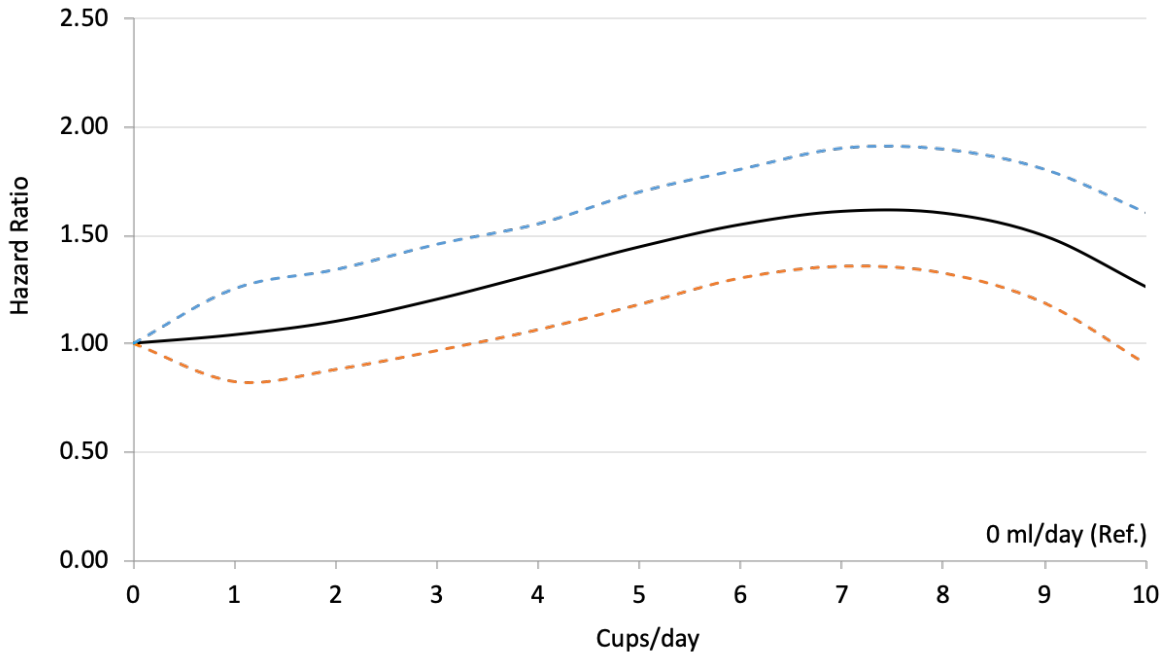
b) Ever Smokers
Non-linear Test: *P*-test=0.07
1 Cup/day Increment Risk: $HR_{-increase}$=1.05, 95% CI=1.04-1.06
P-increase<0.001



c) Never Smokers
Non-linear Test: *P*-test<0.001
1 Cup/day Increment Risk: $HR_{-increase}$=0.99, 95% CI=0.97-1.01
P-increase=0.15



d) Males
Non-linear Test: *P*-test=0.10
1 Cup/day Increment Risk: $HR_{-increase}$=1.06, 95% CI=1.05-1.07
P-increase<0.001



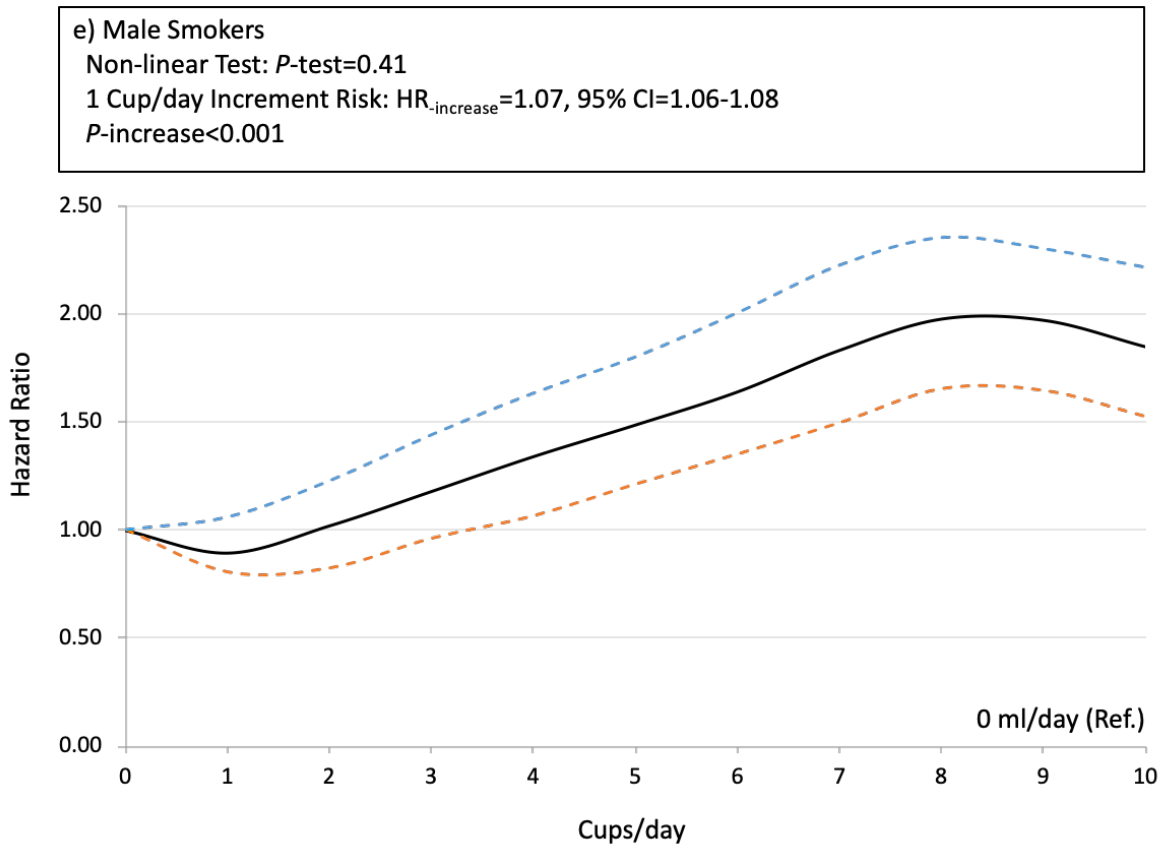


Figure 4-1 *Dose-response Relationships between Coffee Consumption and the Risk of Bladder Cancer among a) Overall Study Population; b) Ever Smokers; c) Never Smokers; d) Males; e) Male Smokers*

The solid lines represent the Hazard Ratios (HRs); the dashed lines represent the 95% confidence intervals (CIs) for the trend; the 1 cup represents 125 ml.

The HRs were adjusted for age (years, continuous), sex (male or female, if applicable), smoking (if applicable, smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year, and no information on pack-years)]), total energy intake (kcal/day, continuous), tea consumption (ml/day, continuous), other caffeine intake (g/day, continuous), and sweeteners intake (g/day, continuous) (model 3).

Abbreviation: ml, millilitre; kcal, kilocalorie; g, gram.

Reference group was never coffee consumers.

P-test <0.05 was considered statistically significant.

P-increase <0.05 was considered statistically significant.

Sensitivity Analyses

Compared to participants who were neither smokers nor coffee consumers (**Supplementary Table 4-2**), both current and former smokers showed an increased bladder cancer risk with the every tertile ordered increment of coffee consumption [current smoker: $HR_{increase}$=1.72, 95% CI=1.66-1.78; former smoker: $HR_{increase}$=1.40, 95% CI=1.34-1.46, (model 3)]. Furthermore, the risk estimates for current smokers were higher than the estimates for former smokers. After removing bladder cancer cases diagnosed within the first 2 years after recruitment, similar increased bladder cancer risks were observed (**Supplementary Table 4-3**). The meta-analyses approach showed similar results, that is a significantly increased bladder cancer risk with higher coffee consumption in overall population 1.19 (95% CI 1.02-1.36) and male smokers 1.30 (95% CI 1.05-1.56).

Discussion

This large multicentric cohort study found a consistently increased bladder cancer risk for high (>500 ml/day) coffee consumption; however, the positive association was restricted to male smokers, while no evidence of an association or increased bladder cancer risk was observed among female and never smokers.

Coffee consumption is strongly associated with smoking due to the compound caffeine in coffee which increases subjective smoking reinforcement, and it appears that these two behaviours often occur at the same time [57-59]. It has been proposed that smoking increases caffeine metabolism, thereby requiring smokers to consume more caffeine to achieve desired stimulant effects [60, 61] and postpone experiencing symptoms of caffeine toxicity [33, 62, 63]. Moreover, this hypothesis is strengthened by experimental studies reporting that the *cytochrome P450 1A2* (CYP1A2) metabolic pathway is upregulated by both caffeine and compounds in tobacco smoke, including nicotine and *polycyclic aromatic hydrocarbons* (PAHs) [64-66], so that the effect of caffeine is potentially weaker among smokers than among never smokers. Several previously conducted observational studies, including our BLEND study on data from case-control studies [67] and extensive meta-analysis [28] confirmed the clear experimental findings, in that a higher bladder cancer risks for coffee consumers was only observed among never smokers. In addition, a recent Finish study showed no association between coffee consumption and bladder cancer risk among male smokers [68]. The present

study, however, showed opposite results; while among smokers an increased bladder cancer risk was observed, no such increased bladder cancer risk was observed among never smokers. One previously conducted prospective cohort study in United States showed a modest increased bladder cancer risk with 1 cup/day increment of coffee consumption among smokers but not never smokers, which is in line with the present study [25]. Since conflicting results were observed among both similar and different research designs, the explanation for these contradictive results remain unclear. It underscores, however, the importance of assessing associations between coffee consumption and bladder cancer by smoking. Therefore, future experimental and large pooled observational study should further explore clear interaction between coffee consumption and smoking in relation to bladder cancer risk, in order to provide definite answers on the influence of coffee consumption and bladder cancer risk. In addition, since bladder cancer is strongly smoking-related, residual confounding by smoking is a particular concern, and statistical adjustment for smoking might only partly separate the risk of bladder cancer due to smoking from a possible risk due to coffee consumption.

The present study and several previously conducted observational studies [25, 28, 69] did not observe an association between coffee consumption and bladder cancer among females. One possible explanation is that females are less likely to develop bladder cancer since the metabolic detoxification of carcinogens (*e.g.* from tobacco or coffee) is stronger in females than in males [70]. However, no mechanism for the influence of coffee consumption on bladder cancer risk in different sexes has been reported.

The mechanisms underlying the associations between coffee consumption and bladder cancer are likely mediated by any of the thousands of specific coffee chemicals present in the beans or produced during roasting [71]. However, in most epidemiological studies, assessment of coffee types is generally limited. Although caffeine showed to suppress the activation of the protein kinase *ataxia-telangiectasia mutated* (ATM) and the phosphorylation of the kinase Chk2, both important for the activation of the tumour suppressor gene P53 [72, 73], subsequent research studies have not confirmed this point. Recent studies based on observations and experiments have shown that coffee may play a role in antioxidant activity [74, 75] or enhance the anti-tumour immune response [76]. The present study, as well as our

study based on data from case-control studies [67], observed increased effects on bladder cancer risk among caffeinated coffee consumers; again, the positive association observed in the present study was restricted to smokers. Notably, however, due to the limited available data on decaffeinated coffee consumption, no detailed analysis could be conducted on this coffee type. Therefore, no definitive conclusion can be made whether it is the caffeine in coffee that increases the bladder cancer risk.

Our dose-response analyses present slightly increasing curves in the overall population and among smokers and males up to 8 cups/day (1,000 ml/day). After 8 cups/day a decrease in bladder cancer risk is observed. Although this observation was probably due to the scarcity of the sample size, it might be caused by the diuretic action of the bladder caused by a large dose of caffeine. A recent review suggests that coffee consumption of over at least 250-300 *micrograms* (mg), *i.e.* 2-3 cups of coffee with 200 ml cup size, have an obvious diuretic action. This profound tolerance to the diuretic effect of caffeine, however, is much diminished in individuals who regularly consume coffee, which might enhance the threshold of sensation of diuretic action to 8 cups [77]. Therefore, the contact time between coffee chemicals and bladder is shortened among those with a higher coffee consumption, thereby, decrease the penetration of coffee consumption into deep layer of bladder.

It is known that cancer might cause behaviour changes prior to diagnosis. It could, therefore, be argued that results obtained from cohort studies overestimate the true effect of coffee consumption on bladder cancer risk if a large proportion of cases are diagnosed within the first 2 years after recruitment. However, after exclusion of cases who were diagnosed within the first 2 years after recruitment, results remain similar.

For the present study, data was collected from 12 cohort studies, thereby yielding a large sample size, allowing to perform detailed and accurate analyses. However, the present study also has some limitations: a) limited information was available on possible risk factors, other than age, sex and smoking, for the development of bladder cancer, such as *body mass index* (BMI), physical activity, *socioeconomic status* (SES), disinfection by-products, arsenic in the drinking water, and occupational exposures to potentially carcinogenic chemicals. Although, adjustments for these factors could have influenced the

results, current literature shows that only a small proportion of bladder cancer cases can be attributed to these factors [72]; b) all events other than bladder cancer were analysed as censored. This was done because of lack of specific information (*e.g.* death) on loss-to-follow-up subject. We were therefore, unable to perform competing risk analyses in addition to the survival analyses; c) although status as well as duration and intensity of smoking were taken into account in our analysis, the adjustment for smoking might still be imperfect due to differences in smoking practices (*e.g.* depth of inhalation or amount of inhalation), or differences in types of smoke exposure [67]. In addition, since smoking is perceived as a socially undesirable behaviour, the use of self-reported questionnaires for smoking status, duration, and intensity might have led to underreporting of the actual smoking habits; d) some additives accompanied with coffee consuming could not be adjusted for, such as non-dairy creamer which has been found to alter the biochemical activities of coffee by interacting with coffee components like polyphenol [78]; e) for most included studies, the exposure variable was assessed by FFQs. Therefore, measurement error and misclassification of study participants in terms of the exposure are unavoidable.

Conclusion

In summary, the present study, with over 500,000 participants and 2,600 bladder cancer cases, observed an increased risk between high (>500 ml/day: around >4 cups/day based on European standard cup size) coffee consumption and bladder cancer among male smokers, while no association between coffee consumption and bladder cancer risk was observed among females and never smokers. The inconsistent results between males and females and the absence of an association in never smokers indicate that no strong evidence for causal links between coffee consumption and the development of bladder cancer, and the positive associations among male smokers is possibly caused by residual confounding of smoking.

References

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. *CA Cancer J Clin* 2017;67(1):7-30. doi:10.3322/caac.21387.

2. Antoni S, Ferlay J, Soerjomataram I, Znaor A, Jemal A, Bray F. Bladder Cancer Incidence and Mortality: A Global Overview and Recent Trends. *Eur Urol* 2017;71(1):96-108. doi:10.1016/j.eururo.2016.06.010.

3. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011;61(2):69-90. doi:10.3322/caac.20107.

4. Ploeg M, Aben KK, Kiemeney LA. The present and future burden of urinary bladder cancer in the world. *World J Urol* 2009;27(3):289-293. doi:10.1007/s00345-009-0383-3.

5. Marugame T, Mizuno S. Comparison of prostate cancer mortality in five countries: France, Italy, Japan, UK and USA from the WHO mortality database (1960–2000). *Jpn J Clin Oncol* 2005;35(11):690-691. doi:10.1093/jjco/hyi185.

6. Stewart B, Wild CP. World cancer report 2014. *International Agency for Research on Cancer, World Health Organization*: WHO Press, 2017.

7. Letašiová S, Medveďová A, Šovčíková A *et al*. Bladder cancer, a review of the environmental risk factors. *Environmental Health* 2012;11(Suppl 1):S11. doi:10.1186/1476-069X-11-S1-S11.

8. Botteman MF, Pashos CL, Redaelli A, Laskin B, Hauser R. The health economics of bladder cancer. *Pharmacoeconomics* 2003;21(18):1315-1330. doi:10.1007/BF03262330.

9. Johansson SL, Cohen SM. Epidemiology and etiology of bladder cancer. Semin Surg Oncol. 1997;13(5):291-8. doi:10.1002/(sici)1098-2388(199709/10)13:5<291::aid-ssu2>3.0.co;2-8.

10. Thun M, Linet MS, Cerhan JR, Haiman CA, Schottenfeld D. Cancer epidemiology and prevention. 3rd ed ed. Oxford University Press; 2017. doi: 10.1093/ije/dyy166.

11. Cole P. Coffee-drinking and cancer of the lower urinary tract. *Lancet* 1971;1(7713):1335-1337. doi:10.1016/S0140-6736(71)91890-3.

12. Nagano J, Kono S, Preston DL, Moriwaki H, Sharp GB, Koyama K et al. Bladder-cancer incidence in relation to vegetable and fruit consumption: A prospective study of atomic-bomb survivors. International journal of cancer. 2000;86(1):132-8. doi:10.1002/(sici)1097-0215(20000401)86:1<132::aid-ijc21>3.0.co;2-m.

13. Pujolar AE, Gonzalez CA, Lopez-Abente G, Errezola M, Izarzugaza I, Nebot M et al. Bladder cancer and coffee consumption in smokers and non-smokers in Spain. International journal of epidemiology. 1993;22(1):38-44. doi:10.1093/ije/22.1.38.

14. Bruemmer B, White E, Vaughan TL, Cheney CL. Fluid intake and the incidence of bladder cancer among middle-aged men and women in a three-county area of western Washington. Nutr Cancer. 1997;29(2):163-8. doi:10.1080/01635589709514619.

15. Demirel F, Cakan M, Yalçınkaya F, Topcuoglu M, Altug U. The association between personal habits and bladder cancer in Turkey. International urology and nephrology. 2008;40(3):643-7. doi:10.1007/s11255-008-9331-1.

16. Kobeissi LH, Yassine IA, Jabbour ME, Moussa MA, Dhaini HR. Urinary bladder cancer risk factors: a Lebanese case- control study. Asian Pacific journal of cancer prevention : APJCP. 2013;14(5):3205-11. doi:10.7314/apjcp.2013.14.5.3205.

17. Sugiyama K, Sugawara Y, Tomata Y, Nishino Y, Fukao A, Tsuji I. The association between coffee consumption and bladder cancer incidence in a pooled analysis of the Miyagi Cohort Study and Ohsaki Cohort Study. *Eur J Cancer Prev* 2017;26(2):125-130. doi:10.1097/cej.0000000000000226.

18. Kurahashi N, Inoue M, Iwasaki M, Sasazuki S, Tsugane S. Coffee, green tea, and caffeine consumption and subsequent risk of bladder cancer in relation to smoking status: a prospective study in Japan. Cancer science. 2009;100(2):284-91. doi:10.1111/j.1349-7006.2008.01027.x.

19. Vena JE, Freudenheim J, Graham S, Marshall J, Zielezny M, Swanson M et al. Coffee, cigarette smoking, and bladder cancer in western New York. Annals of epidemiology. 1993;3(6):586-91. doi:10.1016/1047-2797(93)90079-J.

20. Geoffroy-Perez B, Cordier S. Fluid consumption and the risk of bladder cancer: Results of a multicenter case-control study. International journal of cancer. 2001;93(6):880-7. doi:10.1002/ijc.1411.

21. De Stefani E, Boffetta P, Deneo-Pellegrini H, Correa P, Ronco AL, Brennan P et al. Non-alcoholic beverages and risk of bladder cancer in Uruguay. BMC cancer. 2007;7:57. doi:10.1186/1471-2407-7-57.

22. Mills PK, Beeson WL, Phillips RL, Fraser GE. Bladder cancer in a low risk population: results from the Adventist Health Study. American journal of epidemiology. 1991;133(3):230-9. doi:10.1093/oxfordjournals.aje.a115867.

23. Chyou P-H, Nomura AM, Stemmermann GN. A prospective study of diet, smoking, and lower urinary tract cancer. Annals of epidemiology. 1993;3(3):211-6. doi:10.1016/1047-2797(93)90021-U.

24. Tripathi A, Folsom AR, Anderson KE. Risk factors for urinary bladder carcinoma in postmenopausal women. Cancer. 2002;95(11):2316-23. doi:10.1002/cncr.10975.

25. Loftfield E, Freedman ND, Inoue-Choi M, Graubard BI, Sinha R. A Prospective Investigation of Coffee Drinking and Bladder Cancer Incidence in the United States. *Epidemiology* 2017;28(5):685-693. doi:10.1097/ede.0000000000000676.

26. Loomis D, Guyton KZ, Grosse Y *et al*. Carcinogenicity of drinking coffee, mate, and very hot beverages. *Lancet Oncology* 2016;17(7):877-878. doi:10.1016/S1470-2045(16)30239-X.

27. Swanson JA, Lee JW, Hopp JW. Caffeine and nicotine: a review of their joint use and possible interactive effects in tobacco withdrawal. *Addict Behav* 1994;19(3):229-256. doi:10.1016/0306-4603(94)90027-2.

28. Zhou Y, Tian C, Jia C. A dose-response meta-analysis of coffee consumption and bladder cancer. *Prev Med* 2012;55(1):14-22. doi:10.1016/j.ypmed.2012.04.020.

29. Benowitz NL, Peng M, Jacob P. Effects of cigarette smoking and carbon monoxide on chlorzoxazone and caffeine metabolism. *Clin Pharmacol Ther* 2003;74(5):468-474. doi:10.1016/j.clpt.2003.07.001.

30. Zevin S, Benowitz NL. Drug interactions with tobacco smoking. *Clin Pharmacokinet* 1999;36(6):425-438. doi:10.2165/00003088-199936060-00004.

31. De Leon J, Diaz FJ, Rogers T *et al*. A pilot study of plasma caffeine concentrations in a US sample of smoker and nonsmoker volunteers. *Prog Neuropsychopharmacol Biol Psychiatry* 2003;27(1):165-171. doi:10.1016/S0278-5846(02)00348-2.

32. Plowchalk DR, Yeo KR. Prediction of drug clearance in a smoking population: modeling the impact of variable cigarette consumption on the induction of CYP1A2. *Eur J Clin Pharmacol* 2012;68(6):951-960. doi:10.1007/s00228-011-1189-y.

33. Yu EY, Wesselius A, van Osch F, Stern MC, Jiang X, Kellen E et al. The association between coffee consumption and bladder cancer in the bladder cancer epidemiology and nutritional determinants (BLEND) international pooled study. Cancer causes & control : CCC. 2019. doi:10.1007/s10552-019-01191-1.

34. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol. 1997;26 Suppl 1:S6-14. doi:10.1093/ije/26.suppl_1.s6.

35. Clavel-Chapelon F, van Liere MJ, Giubout C, Niravong MY, Goulard H, Le Corre C et al. E3N, a French cohort study on cancer risk factors. E3N Group. Etude Epidemiologique aupres de

femmes de l'Education Nationale. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP). 1997;6(5):473-8. doi:10.1097/00008469-199710000-00007.

36. Boeing H, Korfmann A, Bergmann MM. Recruitment procedures of EPIC-Germany. European Investigation into Cancer and Nutrition. Annals of nutrition & metabolism. 1999;43(4):205-15. doi:10.1159/000012787.

37. 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. doi:10.1079/phn2002394.

38. Panico S, Iacovo RD, Celentano E, Galasso R, Muti P, Salvatore M et al. Progetto ATENA, a study on the etiology of major chronic diseases in women: design, rationale and objectives. European journal of epidemiology. 1992;8(4):601-8. doi:10.1007/bf00146383.

39. Beulens JW, Monninkhof EM, Verschuren WM, van der Schouw YT, Smit J, Ocke MC et al. Cohort profile: the EPIC-NL study. Int J Epidemiol. 2010;39(5):1170-8. doi:10.1093/ije/dyp217.

40. Lund E, Dumeaux V, Braaten T, Hjartaker A, Engeset D, Skeie G et al. Cohort profile: The Norwegian Women and Cancer Study--NOWAC--Kvinner og kreft. Int J Epidemiol. 2008;37(1):36-41. doi:10.1093/ije/dym137.

41. Manjer J, Carlsson S, Elmstahl S, Gullberg B, Janzon L, Lindstrom M et al. The Malmo Diet and Cancer Study: representativity, cancer incidence and mortality in participants and non-participants. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP). 2001;10(6):489-99. doi:10.1097/00008469-200112000-00003.

42. 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. doi:10.1080/14034950310001432.

43. Davey GK, Spencer EA, Appleby PN, Allen NE, Knox KH, Key TJ. EPIC-Oxford: lifestyle characteristics and nutrient intakes in a cohort of 33 883 meat-eaters and 31 546 non meat-eaters in the UK. Public health nutrition. 2003;6(3):259-69. doi:10.1079/phn2002430.

44. Day N, Oakes S, Luben R, Khaw KT, Bingham S, Welch A et al. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. British journal of cancer. 1999;80 Suppl 1:95-103.

45. van den Brandt PA, Goldbohm RA, van 't Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. Journal of clinical epidemiology. 1990;43(3):285-95. doi:10.1016/0895-4356(90)90009-E.

46. White E, Patterson RE, Kristal AR, Thornquist M, King I, Shattuck AL et al. VITamins And Lifestyle cohort study: study design and characteristics of supplement users. Am J Epidemiol. 2004;159(1):83-93. doi:10.1093/aje/kwh010.

47. Ozasa K, Shimizu Y, Sakata R, Sugiyama H, Grant EJ, Soda M et al. Risk of cancer and non-cancer diseases in the atomic bomb survivors. Radiation protection dosimetry. 2011;146(1-3):272-5. doi:10.1093/rpd/ncr168.

48. Radiation Effects Research Foundation Ethics Committee. https://wwwrerforjp/en/programs/general_research_e/committee/. Accessed 15 July, 2019.

49. Goossens ME, Isa F, Brinkman M *et al*. International pooled study on diet and bladder cancer: the bladder cancer, epidemiology and nutritional determinants (BLEND) study: design and baseline characteristics. *Archives of Public Health* 2016;74(1):30-39. doi:10.1186/s13690-016-0140-1.

50. Satia-Abouta J, Patterson RE, King IB, Stratton KL, Shattuck AL, Kristal AR et al. Reliability and validity of self-report of vitamin and mineral supplement use in the vitamins and lifestyle study. Am J Epidemiol. 2003;157(10):944-54. doi:10.1093/aje/kwg039.

51. Zeegers MP, Goldbohm RA, van den Brandt PA. Are retinol, vitamin C, vitamin E, folate and carotenoids intake associated with bladder cancer risk? Results from the Netherlands Cohort Study. British journal of cancer. 2001;85(7):977-83. doi:10.1054/bjoc.2001.1968.

52. Sauvaget C, Allen N, Hayashi M, Spencer E, Nagano J. Validation of a food frequency questionnaire in the Hiroshima/Nagasaki Life Span Study. Journal of epidemiology. 2002;12(5):394-401. doi:10.2188/jea.12.394.

53. Ferrari P, Slimani N, Ciampi A, Trichopoulou A, Naska A, Lauria C et al. Evaluation of under- and overreporting of energy intake in the 24-hour diet recalls in the European Prospective Investigation into Cancer and Nutrition (EPIC). Public health nutrition. 2002;5(6b):1329-45. doi:10.1079/phn2002409.

54. Poortvliet E, Klensin J, Kohlmeier L. Rationale document for the Eurocode 2 food coding system (version 91/2). *Eur J Clin Nutr* 1992;46(Suppl 5):S9-S24.

55. Orsini N, Li R, Wolk A, Khudyakov P, Spiegelman D. Meta-analysis for linear and nonlinear dose-response relations: examples, an evaluation of approximations, and software. Am J Epidemiol. 2012;175(1):66-73. doi:10.1093/aje/kwr265.

56. Desquilbet L, Mariotti F. Dose-response analyses using restricted cubic spline functions in public health research. Stat Med. 2010;29(9):1037-57. doi:10.1002/sim.3841.

57. Istvan J, Matarazzo JD. Tobacco, alcohol, and caffeine use: a review of their interrelationships. Psychol Bull. 1984;95(2):301-26. doi:10.1037/0033-2909.95.2.301.

58. Emurian HH, Nellis MJ, Brady JV, Ray RL. Event time-series relationship between cigarette smoking and coffee drinking. Addict Behav. 1982;7(4):441-4. doi:10.1016/0306-4603(82)90016-8.

59. Lane JD. Association of coffee drinking with cigarette smoking in the natural environment. Experimental and clinical psychopharmacology. 1996;4(4):409. doi:10.1037/1064-1297.4.4.409.

60. Brown CR, Jacob P, 3rd, Wilson M, Benowitz NL. Changes in rate and pattern of caffeine metabolism after cigarette abstinence. Clinical pharmacology and therapeutics. 1988;43(5):488-91. doi:10.1038/clpt.1988.63.

61. Swanson JA, Lee JW, Hopp JW, Berk LS. The impact of caffeine use on tobacco cessation and withdrawal. Addict Behav. 1997;22(1):55-68. doi:10.1016/S0306-4603(96)00023-8.

62. Langmann P, Bienert A, Zilly M, Väth T, Richter E, Klinker H. Influence of smoking on cotinine and caffeine plasma levels in patients with alcoholic liver cirrhosis. *Eur J Med Res* 2000;5(5):217-221.

63. Ossip DJ, Epstein LH. Relative effects of nicotine and coffee on cigarette smoking. *Addict Behav* 1981;6(1):35-39. doi:10.1016/S0306-4603(81)80006-8.

64. Landi MT, Sinha R, Lang NP, Kadlubar FF. Human cytochrome P4501A2. IARC scientific publications. 1999(148):173-95. doi:PMID: 10493258.

65. Kalow W, Tang BK. Use of caffeine metabolite ratios to explore CYP1A2 and xanthine oxidase activities. Clinical pharmacology and therapeutics. 1991;50(5 Pt 1):508-19. doi:10.1038/clpt.1991.176.

66. Gunes A, Dahl M-L. Variation in CYP1A2 activity and its clinical implications: influence of environmental factors and genetic polymorphisms. *Pharmacogenomics* 2008;9(5):625-637. doi:10.2217/14622416.9.5.625.

67. Yu EY, Wesselius A, van Osch F, Stern MC, Jiang X, Kellen E et al. The association between coffee consumption and bladder cancer in the bladder cancer epidemiology and nutritional determinants (BLEND) international pooled study. Cancer causes & control : CCC. 2019;30(8):859-70. doi:10.1007/s10552-019-01191-1.

68. Hashemian M, Sinha R, Murphy G, Weinstein SJ, Liao LM, Freedman ND et al. Coffee and tea drinking and risk of cancer of the urinary tract in male smokers. Annals of epidemiology. 2019. doi:10.1016/j.annepidem.2019.03.014.

69. Wu W, Tong Y, Zhao Q, Yu G, Wei X, Lu Q. Coffee consumption and bladder cancer: a meta-analysis of observational studies. *Sci Rep* 2015;5:1-9. doi:10.1038/srep09051.

70. Dobruch J, Daneshmand S, Fisch M, Lotan Y, Noon AP, Resnick MJ et al. Gender and Bladder Cancer: A Collaborative Review of Etiology, Biology, and Outcomes. European urology. 2016;69(2):300-10. doi:10.1016/j.eururo.2015.08.037.

71. Rothwell JA, Loftfield E, Wedekind R, Freedman N, Kambanis C, Scalbert A et al. A Metabolomic Study of the Variability of the Chemical Composition of Commonly Consumed Coffee Brews. Metabolites. 2019;9(1). doi:10.3390/metabo9010017.

72. Zhang Z-W, Xiao J, Luo W, Wang B-H, Chen J-M. Caffeine suppresses apoptosis of bladder cancer RT4 cells in response to ionizing radiation by inhibiting ataxia telangiectasia mutated-Chk2-p53 axis. *Chin Med J* 2015;128(21):2938-2945. doi:10.4103/0366-6999.168065.

73. Sarkaria JN, Busby EC, Tibbetts RS, Roos P, Taya Y, Karnitz LM et al. Inhibition of ATM and ATR kinase activities by the radiosensitizing agent, caffeine. Cancer Res. 1999;59(17):4375-82. doi:PMID: 10485486.

74. Tej G, Nayak PK. Mechanistic considerations in chemotherapeutic activity of caffeine. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie. 2018;105:312-9. doi:10.1016/j.biopha.2018.05.144.

75. Azam S, Hadi N, Khan NU, Hadi SM. Antioxidant and prooxidant properties of caffeine, theobromine and xanthine. Med Sci Monit. 2003;9(9):Br325-30. doi:PMID: 12960921.

76. Venkata Charan Tej GN, Neogi K, Verma SS, Chandra Gupta S, Nayak PK. Caffeine-enhanced anti-tumor immune response through decreased expression of PD1 on infiltrated cytotoxic T lymphocytes. European journal of pharmacology. 2019;859:172538. doi:10.1016/j.ejphar.2019.172538.

77. Maughan RJ, Griffin J. Caffeine ingestion and fluid balance: a review. Journal of human nutrition and dietetics : the official journal of the British Dietetic Association. 2003;16(6):411-20. doi:10.1046/j.1365-277x.2003.00477.x.

78. Renouf M, Marmet C, Guy P, Fraering AL, Longet K, Moulin J et al. Nondairy creamer, but not milk, delays the appearance of coffee phenolic acid equivalents in human plasma. The Journal of nutrition. 2010;140(2):259-63. doi:10.3945/jn.109.113027.

Supplementary Materials to Chapter 4

Supplementary Table 4-1 Additional Baseline Characteristic of Included Studies

Supplementary Table 4-2 Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Amount Stratified by Smoking Status

Supplementary Table 4-3 Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Amount Stratified by Sex and Smoking Based on Model 3 [b, c, d] (Removing bladder cancer Cases Diagnosed within the First 2 Years After Recruitment)

Supplementary Figure 4-1 Dose-response Relationships between Coffee Consumption and the Risk of Bladder Cancer among a) Male Never Smokers; b) Females; c) Female Never Smokers; d) Female Smokers

Supplementary Figure 4-2 Forest Plot of Meta-Analyses with HRs and 95% CIs for Highest vs. Never Coffee Consumption with Bladder Cancer Risk in a) Overall Population, b) Male Smokers, c) Male Never Smokers, d) Female Smokers, e) Female Never Smokers

Supplementary Table 4-1 *Additional Baseline Characteristic of Included Studies*

Cohort	No. of Participants (%) [a]	Follow-up Years [b]	Age at Baseline [c]	Person years	No. of Incidences Bladder Cancer	Years to Diagnosis [d]	% of Ever Smoker	Initiate year of baseline assessment
EPIC (Europe)								
France	64,842	10.39	42-71	673,597	31	7.25	0.29	1993
Germany	49,181	9.88	20-70	486,079	216	6.18	0.54	1994
Greece	25,266	9.51	19-84	240,371	50	5.59	0.44	1993
Italy	45,204	11.23	24-78	507,745	192	6.82	0.55	1992
Spain	40,782	12.06	29-70	491,902	154	7.77	0.45	1992
Sweden	49,155	13.12	29-74	644,855	302	8.23	0.51	1991
Netherlands	37,036	11.81	20-70	437,520	119	7.30	0.62	1993
United Kingdom	74,973	11.14	20-98	834,837	250	7.00	0.44	1993
Norway	33,855	9.73	41-56	329,421	24	5.81	0.51	1998
NLCS (Netherlands)	5,243	14.06	55-70	73,646	875	9.34	0.68	1986
VITAL (USA)	69,491	6.75	50-76	468,684	346	3.87	0.64	2000
RERF-LSS (Japan)	6,576	28.04	19-91	184,364	42	22.41	0.46	1960
Overall	501,604	10.71	19-98	5,373,021	2,601	7.70	0.49	1960-2000

Abbreviations: EPIC, European Prospective Investigation into Cancer and Nutrition; NLCS, The Netherlands Cohort Study; VITAL, VITamins And Lifestyle study; RERF-LSS, Radiation Effects Research Foundation-Life Span Study.

[a] As a result of exclusion criteria, the cohort sizes and number of cases included in the pooled analyses may differ from original study-specific publications.

[b] Median years of the time interval from two years after study entry to the date of the last follow-up (e.g. date of death, lost to follow-up, or study exit, whichever came first).

[c] Age ranges at baseline (minimum to maximum).

[d] Time-to-bladder cancer defined by median years of study entry to the date of diagnosis among bladder cancer cases.

Supplementary Table 4-2 *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Amount Stratified by Smoking Status*

Models		Model 1 [e]	Model 2 [e, f]	Model 3 [e, f, g]
Ref.		Ref.	Ref.	Ref.
Never smokers (ml/day)	≤180	0.81 (0.62-1.06)	0.91 (0.69-1.18)	0.93 (0.71-1.21)
	180-500	1.70 (1.34-2.14)	1.47 (1.17-1.84)	1.43 (1.13-1.81)
	>500	1.37 (1.05-1.79)	1.21 (0.92-1.58)	1.23 (0.94-1.61)
	HR (95% CI) Per tertile increase	1.14 (1.07-2.08)	1.10 (1.03-1.17)	1.10 (1.03-1.17)
Current smokers (ml/day)	≤ 180	3.52 (2.74-4.51)	3.42 (2.66-4.38)	3.50 (2.72-4.49)
	180-500	7.88 (6.37-9.74)	5.65 (4.56-6.99)	5.40 (4.32-6.76)
	> 500	8.62 (7.06-10.53)	6.78 (5.55-8.29)	6.49 (5.24-8.03)
	HR (95% CI) Per tertile increase	1.83 (1.77-1.88)	1.73 (1.67-1.78)	1.72 (1.66-1.78)
Former smokers (ml/day)	≤ 180	2.84 (2.24-3.61)	1.66 (1.31-2.11)	1.65 (1.29-2.11)
	180-500	6.01 (4.92-7.34)	2.75 (2.24-3.37)	2.64 (2.13-3.27)
	> 500	6.32 (5.16-7.73)	3.04 (2.48-3.73)	2.91 (2.34-3.61)
	HR (95% CI) Per tertile increase	1.72 (1.67-1.77)	1.42 (1.36-1.47)	1.40 (1.34-1.46)

[e] Model 1: Adjusted for age (years, continuous), sex (male or female).
[e, f] Model 2: Additionally, adjusted for smoking (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]).
[e, f, g] Model 3: Additionally, adjusted for total energy intake (kcal/day, continuous), tea consumption (ml/day, continuous), other caffeine intake (g/day, continuous), and sweeteners intake (g/day, continuous).
Abbreviation: HR, hazard ratio; CI, confidence interval; g, gram; ml, millilitre; kcal, kilocalorie.
Reference group was participants who were neither smokers nor coffee consumers.
P-trend <0.05 was considered statistically significant.

Supplementary Table 4-3 *Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Amount Stratified by Sex and Smoking Based on Model 3 [b, c, d] (Removing Bladder Cancer Cases Diagnosed within the First 2 Years After Recruitment)*

Study Subgroup		Overall	Male				
			Current Smoker	Former Smoker	Never Smoker	Heavy Smoker [h]	Light Smoker [h]
No. Cases/Total Participants		2,289/501,292	695/38,435	783/ 61,886	218/54,579	626/17,698	220/11,753
	Never	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
	≤180 ml/day	1.10 (0.94-1.28)	1.21 (0.88-1.67)	1.12 (0.83-1.50)	0.74 (0.47-1.15)	1.13 (0.76-1.67)	1.44 (0.81-2.57)
	180-500 ml/day	1.46 (1.27-1.67)	1.51 (1.07-2.12)	1.25 (0.95-1.65)	0.96 (0.64-1.45)	1.56 (1.03-2.37)	1.45 (0.80-2.62)
	>500 ml/day	1.70 (1.49-1.95)	1.54 (1.10-2.17)	1.46 1.11-1.92)	0.86 (0.56-1.33)	1.74 (1.14-2.65)	1.82 (1.08-3.22)
P-trend		<0.001	0.02	0.01	0.71	0.01	0.04

Supplementary Table 4-3_continued *Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Coffee Consumption Amount Stratified by Sex and Smoking Based on Model 3 [b, c, d] (Removing Bladder Cancer Cases Diagnosed within the First 2 Years After Recruitment)*

Study Subgroup		Female				
		Current Smoker	Former Smoker	Never Smoker	Heavy Smoker [h]	Light Smoker [h]
No. Cases/Total Participants		178/ 61,467	164/83,293	251/201,632	106/20,119	84/35,304
	Never	Ref.	Ref.	Ref.	Ref.	Ref.
	≤180 ml/day	0.80 (0.46-1.38)	0.68 (0.38-1.20)	0.94 (0.63-1.40)	0.91 (0.42-1.97)	0.59 (0.27-1.31)
	180-500 ml/day	0.76 (0.44-1.30)	0.78 (0.47-1.29)	0.95 (0.64-1.41)	0.93 (0.50-2.06)	0.39 (0.17-1.02)
	>500 ml/day	0.88 (0.52-1.49)	0.80 (0.50-1.30)	0.90 (0.58-1.38)	0.87 (0.41-1.82)	0.67 (0.30-1.50)
P-trend		0.80	0.57	0.65	0.62	0.66

[e, f, g] Model 3: Adjusted for age (years, continuous), sex (male or female), smoking (if applicable, smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), for total energy intake (kcal/day, continuous), tea consumption (ml/day, continuous), other caffeine intake (g/day, continuous), and sweeteners intake (g/day, continuous).

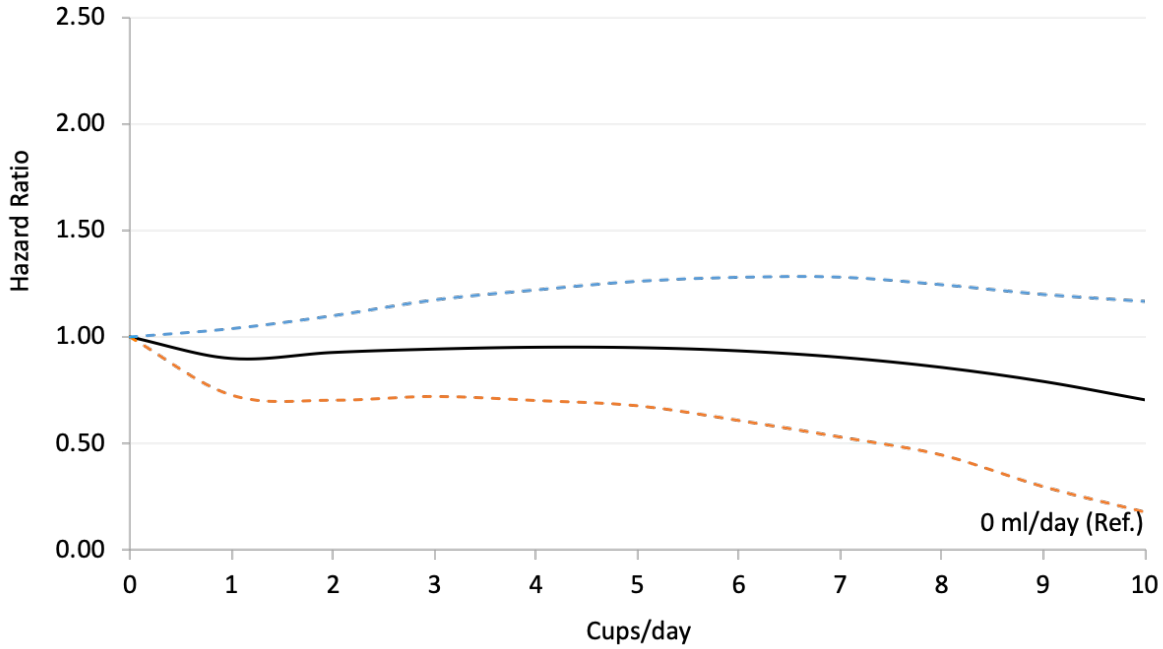
Abbreviation: g, gram; ml, millilitre; kcal, kilocalorie.

[h] Heavy Smoker: Pack-years >20; Light Smoker: Pack-years ≤20; and both heavy and light smokers included current and former smokers.

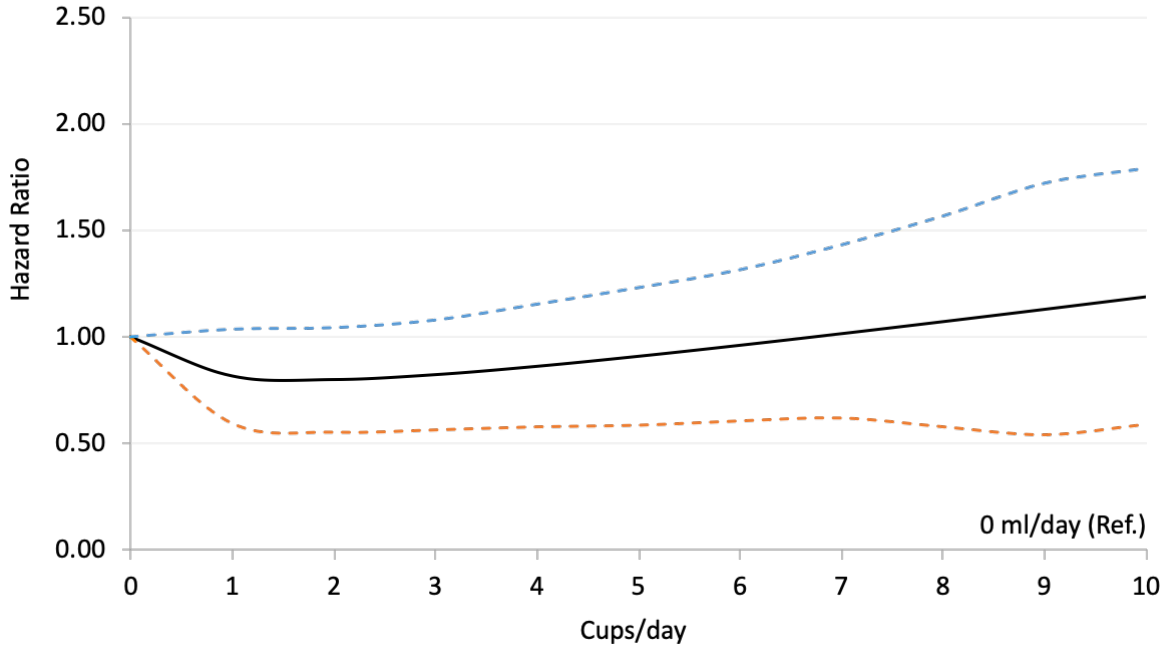
Referent group was never coffee consumers.

P-trends <0.05 was considered statistically significant.

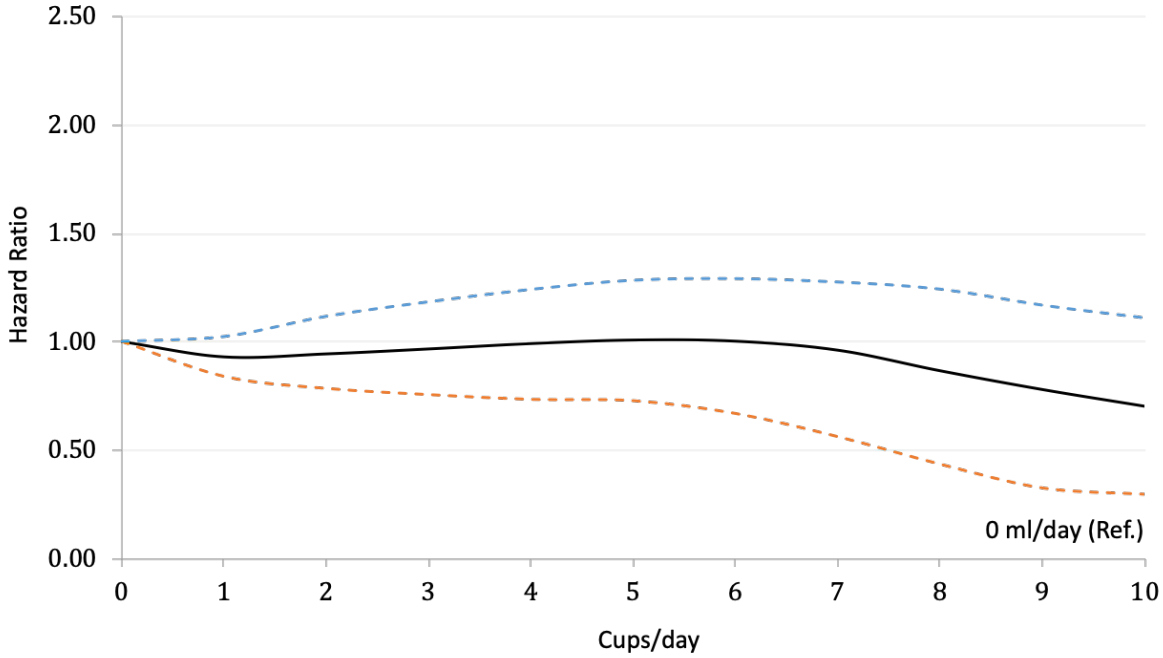
a) Male Never Smokers
Non-linear Test: *P*-test=0.01
1 Cup/day Increment Risk: $HR_{-increase}$=0.99, 95% CI=0.96-1.02
P-increase=0.10



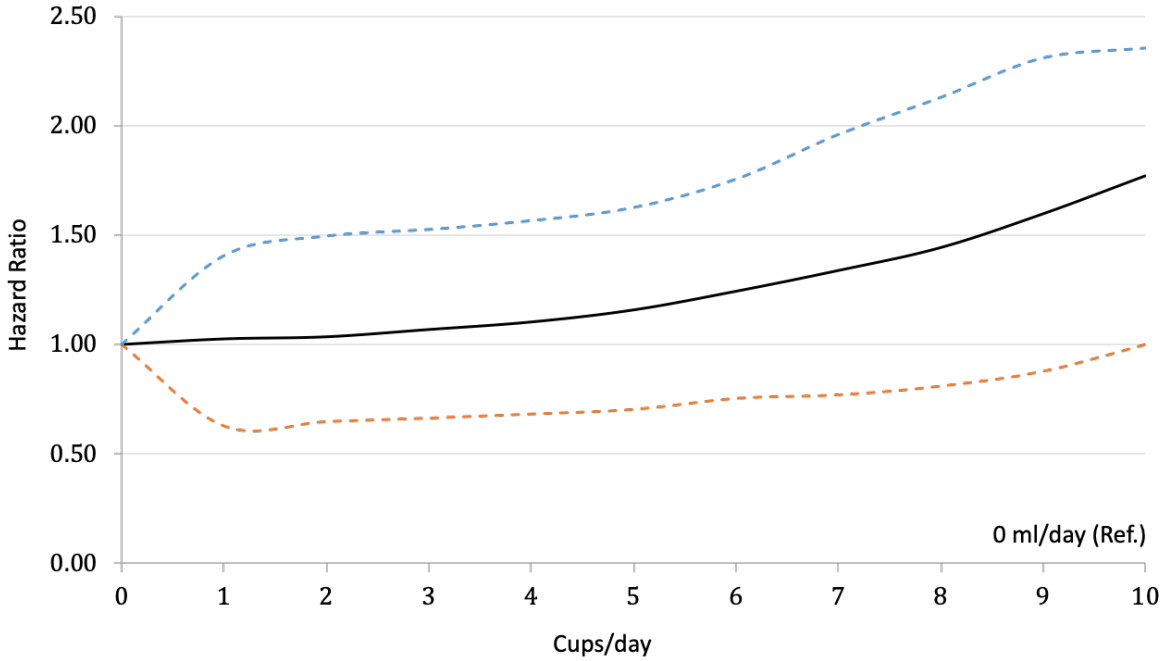
b) Females
Non-linear Test: *P*-test<0.001
1 Cup/day Increment Risk: $HR_{-increase}$=1.02, 95% CI = 0.97-1.07
P-increase=0.62



c) Female Never Smokers
Non-linear Test: *P*-test=0.01
1 Cup/day Increment Risk: $HR_{-increase}$=0.98, 95% CI=0.96-1.00
P-increase=0.13



d) Female Ever Smokers
Non-linear Test: *P*-test<0.001
1 Cup/day Increment Risk: $HR_{-increase}$=1.05, 95% CI=0.99-1.11
P-increase=0.11



Supplementary Figure 4-1 *Dose-response Relationships between Coffee Consumption and the Risk of Bladder Cancer among a) Male Never Smokers; b) Females; c) Female Never Smokers; d) Female Smokers*

The solid lines represent the hazard ratios (HRs); the dashed lines represent the 95% confidence intervals (CIs) for the trend; the 1 cup represents 125 ml.

The HRs were adjusted for age (years, continuous), sex (male or female), smoking (if applicable, smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), total energy intake (kcal/day, continuous), tea consumption (ml/day, continuous), other caffeine intake (g/day, continuous), and sweeteners intake (g/day, continuous) (model 3).

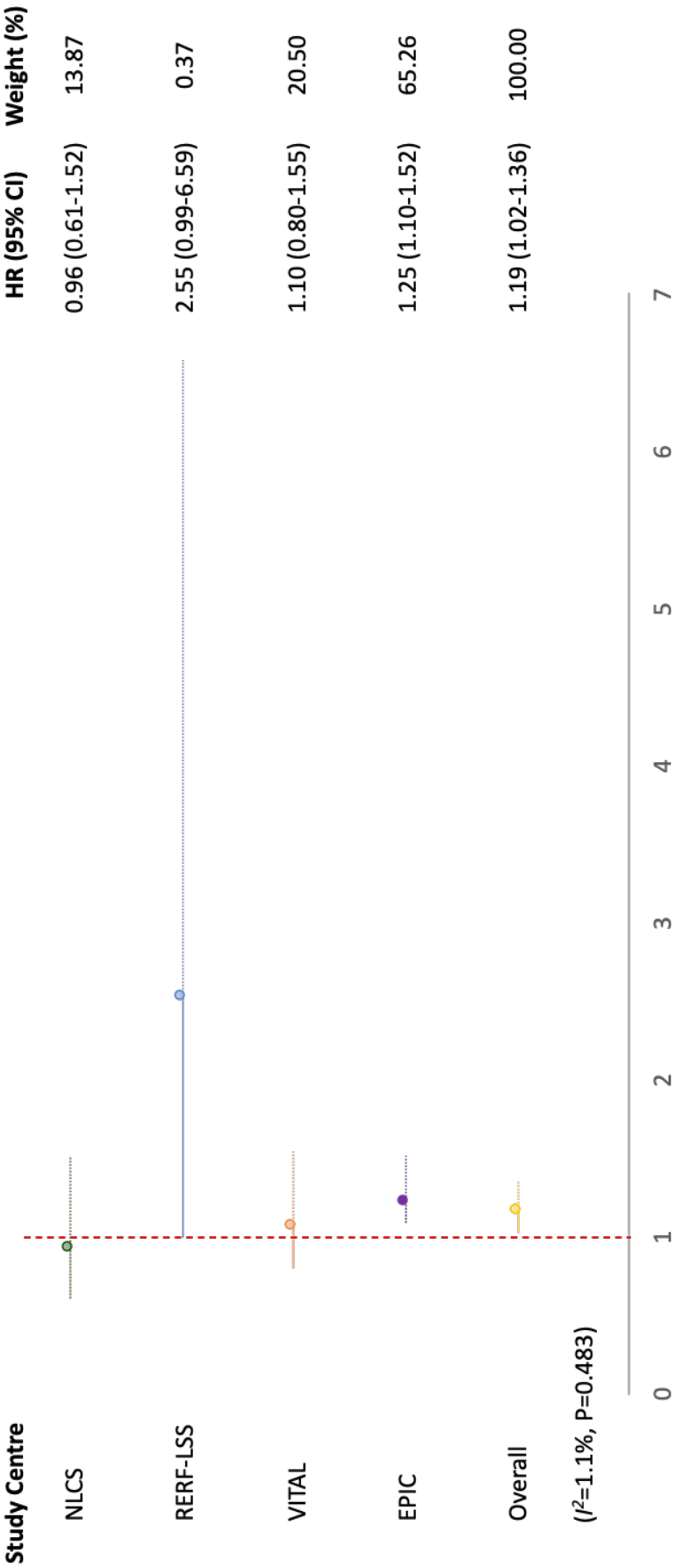
Abbreviation: ml, millilitre; kcal, kilocalorie; g, gram.

Reference group were never coffee consumers.

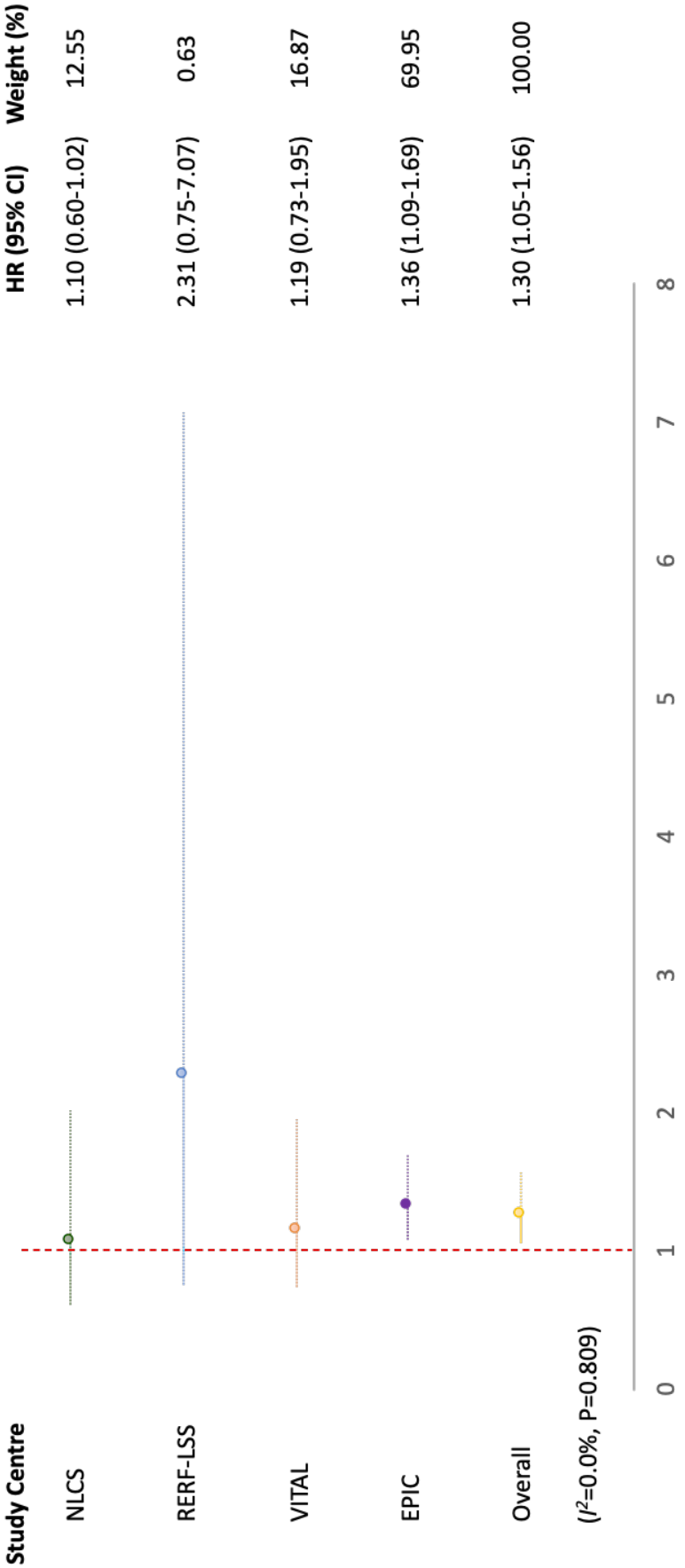
P-test <0.05 was considered statistically significant.

P-increase <0.05 was considered statistically significant.

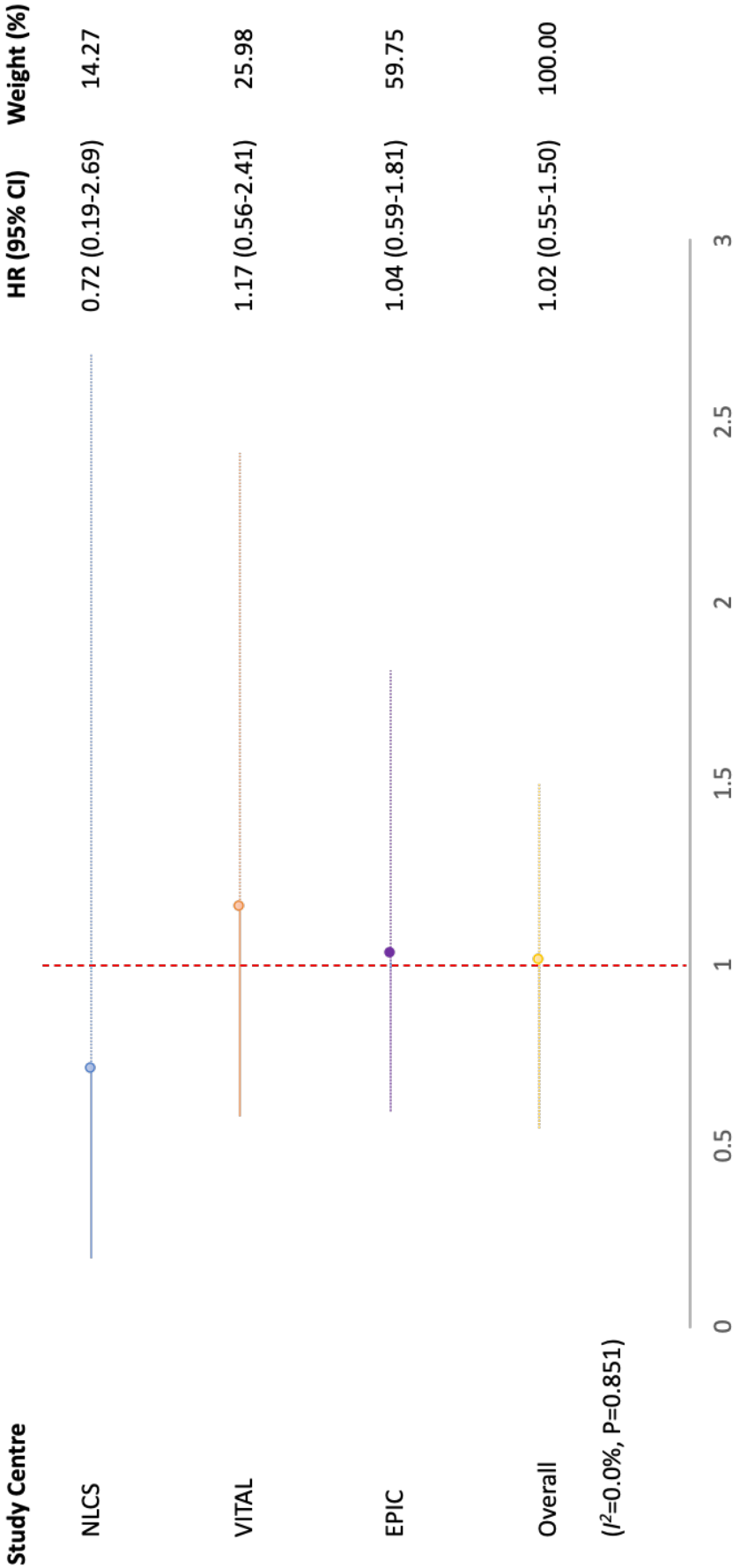
a) Overall Population



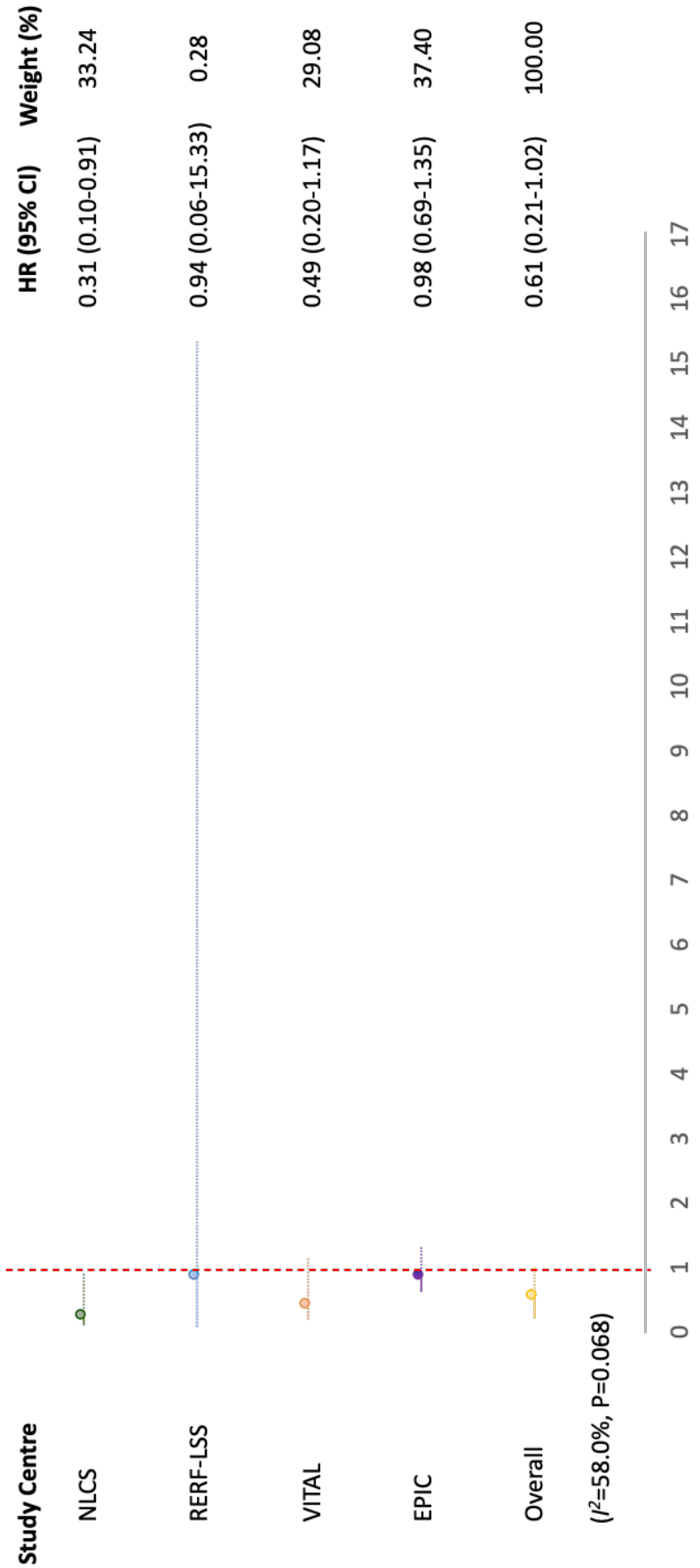
b) Male Smokers



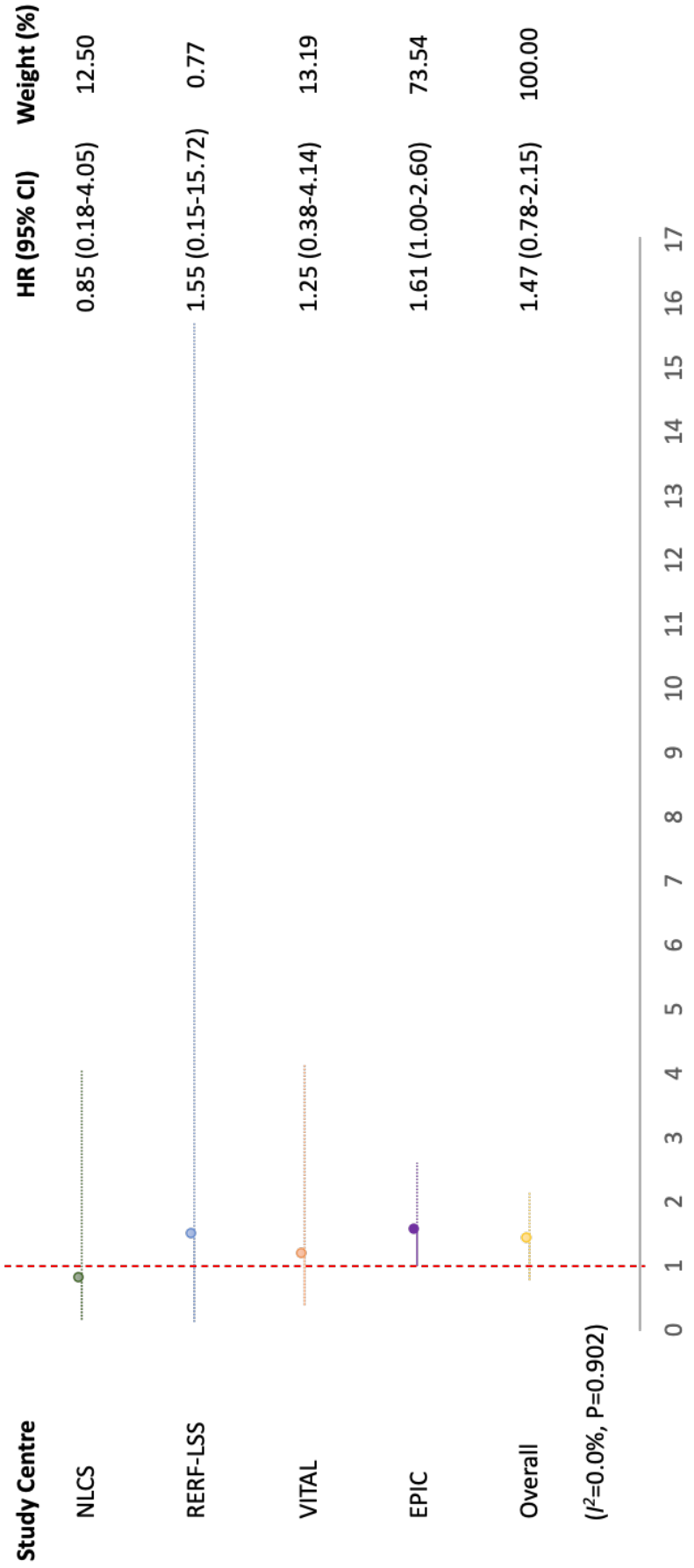
c) Male Never Smokers



d) Female Smokers



e) Female Never Smokers



Supplementary Figure 4-2 *Forest Plot of Meta-Analyses with HRs and 95% CIs for Highest vs. Never Coffee Consumption with Bladder Cancer Risk in a) Overall Population, b) Male Smokers, c) Male Never Smokers, d) Female Smokers, e) Female Never Smokers*

Diamond dots denote the hazard ratios (HRs); Horizontal lines represent the 95% confidence intervals (CIs); Weights (grey squares) are from random effects analyses.

The HRs were adjusted for age (years, continuous), sex (male or female), smoking (if applicable, smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), total energy intake (kcal/day, continuous), tea consumption (ml/day, continuous), other caffeine intake (g/day, continuous), and sweeteners intake (g/day, continuous) (model 3).

Abbreviation: ml, millilitre; kcal, kilocalorie; g, gram.

Reference group were never coffee consumers.

P-value <0.05 was considered statistically heterogeneous.

Chapter 5

Vegetable Intake and Bladder Cancer Risk: A Pooled Analysis of 580,953 Participants from 12 Prospective Cohort Studies

Evan Y.W. Yu, Anke Wesselius, Siamak Mehrkanoon, Mieke Goosens, Maree Brinkman, Piet van den Brandt, Eric J. Grant, Emily White, Elisabete Weiderpass, Florence Le Calvez-Kelm, Marc Gunter, Inge Huybrechts, Fredrik Liedberg, Marko Lukic, Anne Tjonneland, Elio Riboli, Graham G. Giles, Roger L. Milne, Maurice P. Zeegers

Abstract

Background: Although a potential protective effect of vegetable intake on bladder cancer risk has been reported for many years, epidemiological evidence is inconsistent. This research aims to elucidate the association between vegetable intake and bladder cancer risk by bringing together worldwide cohort studies on this topic.

Methods: Vegetable intake in relation to bladder cancer risk was examined by pooling individual data from 14 cohort studies, comprising of 3,253 cases out of 580,953 participants. Pooled multivariate *hazard ratios* (HRs), with corresponding 95% *confidence intervals* (CIs), were estimated using Cox regression models stratified on cohort for intakes of total vegetables, vegetable subtypes (*i.e*. non-starchy, starchy, green leafy and cruciferous vegetables) and individual vegetable types. Analyses were stratified by sex, smoking status, and bladder cancer subtype.

Results: The association between vegetable intake and bladder cancer risk showed interaction with sex (*P*-interaction=0.013) and smoking status (*P*-interaction=0.038). Therefore, analyses were stratified by sex and smoking status. After adjustment for potential confounders, we found that higher intake of total and non-starchy vegetables are inversely associated with the risk of bladder cancer (comparing the highest with lowest intake tertile: HR=0.79, 95% CI=0.64-0.98, P=0.034 for trend; HR per 1 SD increment=0.89, 95% CI=0.81-0.99; HR_{model2}=0.79, 95% CI=0.64-0.98, P=0.037 for trend; HR per 1 SD increment=0.89, 95% CI=0.81-0.99, respectively) among females. In addition, a decreased bladder cancer risk was observed for high intake of curly kale (comparing the highest with lowest intake tertile: HR_{model2} =0.80, 95% CI=0.68-0.95; comparing the medium with lowest intake tertile: HR_{model2} =0.86, 95% CI=0.73-1.00, P=0.072 for trend; HR per 1 SD increment=0.94, 95% CI=0.87-1.01).

Conclusion: Higher intakes of total and non-starchy vegetable are associated with reduced risk of bladder cancer for females but not for males. Further studies are needed to clarify the underlying mechanisms for these findings.

Key Words: Bladder Cancer; Vegetable; Dietary Diversity Analysis; Cohort Study

Introduction

Bladder cancer is the most common malignancy of urinary tract, with an estimated 550,000 new cases and 200,000 deaths annually [1-3]. Incidence rates of bladder cancer are highest in Europe and North America, with a strong predominance in males and the elderly [4-9]. Due to its high rate of recurrence, bladder cancer is reported to be the most expensive of all cancer in terms of life time treatment [10]. Apart from well-established risk factors, such as smoking, occupational exposure and arsenic contaminant [11-15], dietary factors may also contribute to bladder cancer risk considering that many dietary compounds are excreted in urine and thereby coming into direct contact with the surface of the bladder [12]. As recommended in numerous dietary guidelines globally, intake of vegetable may be beneficial for cancer prevention according to its rich content of vitamins, fibre, minerals, and other bioactive compounds [16-18]. However, in the World Cancer Research Fund (WCRF) International's *Continuous Update Project* (CUP) 2018 [19], the evidence suggesting that higher intake of vegetable decreases the risk of bladder cancer is limited.

Previous epidemiologic studies on the influence on vegetable intake and bladder cancer risk have yielded inconsistent results; While most studies showed null association between total vegetable intake and bladder cancer risk, some showed opposite results, in a higher intake reduces the risk of bladder cancer [20-24], or that a higher intake increases the risk of bladder cancer [25]. Moreover, meta-analyses on this topic also show contrasting results. Meta-analysis based on cohort studies only, found no association between vegetable intake and bladder cancer risk [26], while meta-analysis combining results from case-control and cohort studies showed a protective effect of vegetable intake on bladder cancer [27-29].

In addition, it remains unclear which vegetable subtype (*e.g.* non-starchy, starchy, green leafy and cruciferous) or individual vegetable type (*e.g.* broccoli, garlic, cauliflower, kale, cabbage, asparagus) may be responsible for a possible inverse association with bladder cancer. The potential protective effect could, therefore, potentially be the result of a cumulative influence of many preventative agents from total vegetables intake, or from a anticarcinogenic property of a agents from a specific vegetables (*i.e.* isothiocyanates from broccoli [30-32]; S-allylmercaptocysteine (SAMC) [33] and diallyl trisulfide (DATS) from [34] from garlic).

This inconsistent evidence and lack of knowledge on which vegetable subtypes is possibly causing an effect on bladder cancer risk, could be due to the small sample sizes of previous studies and their consequent lack of statistical power to detect weak associations and perform subgroup analysis [27]. Therefore, this study aimed to investigate and find more conclusive answers on the associations between intake of total vegetable and vegetable subtypes and bladder cancer risk by pooling data from 580,953 participants with 3,253 bladder cancer cases in 14 cohort studies included in the *BLadder Cancer Epidemiology and Nutritional Determinants* (BLEND) international consortium.

Methods

Study Sample

Data were obtained from BLEND, an international nutritional consortium currently consisting of 19 case-control studies and 16 cohort studies. Fourteen cohort studies with a total of 580,953 participants, 3,253 of whom developed incident bladder cancer, had sufficient information on vegetable intake to be eligible for inclusion in the present study. These studies originated from 13 countries in 4 continents [*i.e.* Europe: *European Prospective Investigation into Cancer and Nutrition* cohort study (EPIC) [35] (Denmark [36], France [37], Germany [38], Greece [39], Italy [40], Spain [39], Sweden [41, 42], The Netherlands [43], United Kingdom [44, 45], Norway [46]), *NetherLands Cohort Study* (NLCS) [47]; North America: *VITamins And Lifestyle* cohort study (VITAL) [48]; Oceania: *Melbourne Collaborative Cohort Study* (MCCS) in Australia [49, 50]; and Asia: *Radiation Effects Research Foundation-Life Span Study* (RERF-LSS) [51]]. Person-years of follow-up for each participant were calculated from the date of study enrolment until the date of bladder cancer diagnosis, or date of last follow-up (*e.g.* date of death, lost to follow-up, or study exit), whichever came first. For the NLCS study, a nested case-cohort design was applied in order to increase the follow-up coverage and efficiency, in which the number of person-years at risk was estimated based on a sub-cohort that was randomly sampled [47]. Each study was approved by their local ethical research committee [35, 47, 48, 50, 51] (**Supplementary Table 5-1**).

Data Collection and Coding

Details on the methodology of the BLEND consortium have been described elsewhere [52]. In brief, all included studies used self-administered or trained interviewer administered *food*

frequency questionnaire (FFQ) that was validated on either food groups [48, 53-56], and/or energy intake [53, 56, 57]. For each study, participants were asked to report on their usual vegetable intake during the year before study enrolment. This data was harmonized and categorized in individual vegetable subtype intake using the hierarchal Eurocode 2 food coding system developed by the European Union [58], with weekly, monthly or yearly intake converted to grams (g) per day. The classification of the Eurocode 2 Food Coding System defines which food items are covered for what food category so that the potential for misidentification of foods will be limited.

Each study ascertained incident bladder cancer, defined to include all urinary bladder neoplasms according to the *International Classification of Diseases for Oncology* (ICD-O-3 code C67) using population-based cancer registries, health insurance records, or medical records. Bladder cancers were classified as non-muscle invasive bladder cancer (NMIBC) or muscle invasive bladder cancer (MIBC). NMIBC included non-invasive papillary carcinomas confined to the urothelium (stage Ta), and carcinomas that invaded the lamina propria of the bladder wall (stage T1). High grade flat non-invasive carcinomas confined to the urothelium (carcinoma in situ; CIS) without other concomitant tumour stages [*i.e.* T1/Ta (classified to non-muscle invasive prior), or muscle invasive] were also classified as NMIBC. MIBC included carcinomas that invaded into the detrusor muscle (stage T2), carcinomas that invaded into the peri-vesical tissue (stage T3), and carcinomas that invaded adjacent tissues and organs (most often the prostate or uterus, stage T4).

In addition to information on vegetable and other dietary intakes, the BLEND dataset also included data on study characteristics (design, method of dietary assessment, geographical region), participant demographics (age, sex and ethnicity), smoking status and smoking pack-years (*i.e.* the number of cigarettes smoked per day multiplying the years of smoking), which were measured at baseline.

Statistical Analyses

To assess the influence of vegetable intake on bladder cancer risk, Cox regression analyses with a stratification approach to adjust for cross-cohort heterogeneity [59], were used to estimate the pooled *hazard ratios* (HRs) and 95% *confidence intervals* (CIs). The proportional

hazard assumption was examined for each analysis and no evidence of violation was found. Additionally, the appropriateness of the use of the log-normal distribution was tested using a Wald test, and again, no evidence of violation was found. Total vegetable intake, intake of the main vegetable subgroups [*i.e.* non-starchy (including 3,253 cases/580,953), starchy (including 1,514 cases/96,108 participants), green leafy (including 3,212cases/574,381 participants) and cruciferous (including 3,212/574,381 participants) were expressed as tertiles: low intake (tertile 1), medium intake (tertile 2) and high intake (tertile 3). Low intake was used as the reference group and associations were assessed applying two models. Model 1 was adjusted for age (years), sex (male and female), smoking, total energy intake [*kilocalorie* (kcal)/day, continuous; using residual model to remove extraneous variation [60]]. Model 2 was additionally adjusted for ethnicity (Caucasian or non-Caucasian), alcohol intake [*millilitre* (ml)/day, continuous], sugar intake [gram (g)/day, continuous], meat intake (g/day, continuous), grain intake (g/day, continuous), and fat intake (g/day, continuous), which have been shown to be potential dietary risk factors for bladder cancer risk [61]. Smoking was defined as a dummy variable: 0 (never smokers); 1 [current light smokers (*i.e.* smoking less than 20 pack-years)]; 2 [current heavy smokers (*i.e.* smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (*i.e.* smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (*i.e.* smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]. To understand the relevance of effect modification, the main interaction terms between vegetable intake and age, sex and smoking status were added to model 1. P-interaction <0.05 was considered statistically significant where upon all analyses were stratified for the covariate of interest.

In addition, the HRs and 95% CIs of bladder cancer per 1 *standard deviation* (SD) increase in total vegetables and vegetable subgroups intake were estimated using the same models. To test for linearity or nonlinearity, we included both linear and quadratic terms (*i.e.* the absolute intake and intake squared) in the models and then used a likelihood ratio test to assess the difference between the nonlinear and linear models [62]. Since results showed no evidence of a non-linear association, linear models were applied. P for trend test was conducted by assigning medians to per 1 SD increase in intake as a continuous variable in the models.

The variables of bladder cancer status (*i.e.* cases or non-cases), follow-up time, age, sex, smoking and total energy intake were complete without missing values. Missing values of variables (*e.g.* categorical: ethnicity; continuous: alcohol intake, sugar intake, meat intake, vegetable intake, and fat intake; missing proportion were all less than 5%) were imputed by multiple imputation method. Since study protocols and characteristics of participants across cohorts might differ, multiple imputation was done separately for each included cohort study. Only participants with complete information on bladder cancer status, age, sex, smoking and total energy intake were included in building of the imputation models. Linear regression models were then fitted for those variables with missing data separately.

In our secondary analysis, a *diet diversity score* (DDS) was used to quantify the variety of vegetables intake [52, 53] and to assess the effect of the varied individual types of vegetable intake on bladder cancer risk. DDS was calculated as the total number of different individual vegetable types consumed at least once per week and was then divided in 4 groups; 0-2 low diversity, 2-4 median diversity, 4-6 high diversity, 6-8 very high diversity. To analyse the effect of variety in vegetable intake independent from the amount of vegetables intake, we adjusted the DDS for the total vegetable intake.

A sensitivity analysis was performed by removing cases diagnosed within the first 2 years after recruitment to each study. In addition, the association between intake of total vegetable and risk of bladder cancer was assessed in each participating cohort separately and combined in a meta-analysis approach using a random-effect model.

All statistical analyses were performed using STATA version 14 SE (Stata Corporation, Texas, USA). A two-tailed P values <0.05 was considered statistically significant.

Results

Baseline characteristics of the included study samples are presented in **Table 5-1**. In total, 580,953 study participants contributed 6,505,279 person-years of follow-up over a median of 11 years, with 3,253 incident bladder cancer cases (2,445 males, 808 females) diagnosed. Of these, 2,040 (63%) cases had available diagnosis records of NMIBC (39%) or MIBC (24%). The

median age at baseline was 60 years for cases and 53 for non-cases. The majority (98.2%) of participants were Caucasian. Higher intake of total vegetable, non-starchy vegetable, green leafy vegetable and cruciferous vegetable were observed in bladder cancer cases (P-value <0.001, <0.001, 0.519 and <0.001 respectively), while only higher intake of starchy vegetable was observed in non-bladder-cancer cases (P-value< 0.001).

Associations of Vegetables Intake with Bladder Cancer Risk

Total vegetable intake and bladder cancer risk

For the total vegetable intake, no evidence of association was observed for the overall population and neither for the bladder cancer subtypes (overall: $HR_{model2;\ highest\ vs\ lowest\ tertile}$=0.95, 95% CI=0.85-1.07, P=0.296 for trend, HR per 1 SD increment=1.00, 95% CI=0.94-1.05; NMIBC: $HR_{model2;\ highest\ vs\ lowest\ tertile}$=1.01, 95% CI=0.84-1.21, P=0.404 for trend, HR per 1 SD increment=0.97, 95% CI=0.89-1.05; MIBC: $HR_{model2;\ highest\ vs\ lowest\ tertile}$=0.99, 95% CI=0.77-1.27, P=0.582 for trend, HR per 1 SD increment=0.97, 95% CI=0.97-1.07) (**Table 5-2**). Sex stratified analysis showed an inverse association between bladder cancer risk and high total vegetable intake among females ($HR_{model2;\ highest\ vs\ lowest\ tertile}$ =0.79, 95% CI=0.64-0.98, P=0.034 for trend; HR per 1 SD increment=0.89, 95% CI=0.81-0.99), while no association was observed among males ($HR_{model2;\ highest\ vs\ lowest\ tertile}$=1.02, 95% CI=0.90-1.17, P=0.911 for trend; HR per 1 SD increment=1.00, 95% CI=0.94-1.05). Stratification for smoking only showed a marginally non-significant decreased bladder cancer risk per 1 SD increment of total vegetable intake among current smokers ($HR_{model2;\ highest\ vs\ lowest\ tertile}$=0.94, 95% CI=0.78-1.13, P=0.911 for trend; HR per 1 SD increment=0.91, 95% CI=0.84-0.99). No such decline in risk could be observed among never or former smokers (respectively, $HR_{model2;\ highest\ vs\ lowest\ tertile}$=1.07, 95% CI=0.84-1.36, P=0.795 for trend; HR per 1 SD increment=1.01, 95% CI=0.92-1.10), and former smokers ($HR_{model2;\ highest\ vs\ lowest\ tertile}$=0.92, 95% CI=0.77-1.10, P=0.911 for trend; HR per 1 SD increment=1.00, 95% CI=0.94-1.11).

Table 5-1 *Characteristics of the Study Population (3,253 cases and 577,700 non-cases) and Vegetable Intake of Included Studies*

Characteristics		N (%)					Person-years
			Case (%)	NMIBC (%)	MIBC (%)	Non-case (%)	
	Total	580,953	3,253	1,267	773	577,700	6,516,959
Vegetable Intake (g/day)	Low (0-117)	193,651 (33.33)	997 (30.65)	392 (30.94)	167 (21.60)	192,654 (33.35)	2,104,394
	Medium (117-221)	193,651 (33.33)	971 (29.85)	390 (30.78)	213 (27.55)	192,680 (33.35)	2,155,133
	High (>221)	193,651 (33.33)	1,285 (39.50)	485 (38.28)	393 (50.84)	192,366 (33.30)	2,257,433
	Mean [g/day (SD)]	197.50 (146.00)	207.37 (143.69)	202.11 (131.77)	238.49 (138.60)	197.44 (146.01)	197.50 (146.00)
P [a]		<0.001					<0.001
P [b]							

Table 5-1_continued *Characteristics of the Study Population (3,253 cases and 577,700 non-cases) and Vegetable Intake of Included Studies*

Characteristics		**Sex**				**Age [Mean (SD)]**		**Smoking Pack-years [Mean (SD)]** [c]
		Male (%)		Female (%)				
		Case	Non-case	Case	Non-case	Case	Non-case	
	Total	2,445	188,728	808	388,972	3,253	577,700	147,303
	Low (0-117)	742 (30.35)	81,193 (43.02)	255 (31.56)	111,461 (28.66)	60.01 (8.42)	53.36 (10.40)	25.01 (19.92)
Vegetable Intake (g/day)	Medium (117-221)	688 (28.14)	58,003 (30.73)	283 (35.02)	134,677 (34.62)	60.01 (7.39)	52.31 (9.90)	22.44 (17.69)
	High (>221)	1,015 (41.51)	49,532 (26.25)	270 (33.42)	142,834 (36.72)	60.86 (6.86)	51.95 (10.35)	22.38 (19.21)
	Mean [g/day (SD)]	212.72 (148.17)	175.30 (147.18)	191.17 (127.91)	208.19 (144.22)	207.37 (143.69)	197.44 (146.01)	194.75 (145.85)
P [a]		<0.001		0.090		<0.001		<0.001
P [b]						0.466		

Table 5.1_continued *Characteristics of the Study Population (3,253 cases and 577,700 non-cases) and Vegetable Intake of Included Studies*

Characteristics		Smoking Status					
		Never (%)		Current (%)		Former (%)	
		Case	Non-case	Case	Non-case	Case	Non-case
	Total	663	289,141	1,224	120,443	1,366	168,116
Vegetable Intake (g/day)	Low (0-117)	196 (29.56)	87,195 (30.16)	391 (31.94)	46,322 (38.46)	410 (30.01)	59,137 (35.18)
	Medium (117-221)	209 (31.52)	95,177 (32.92)	360 (29.41)	40,836 (33.90)	402 (29.43)	56,667 (33.71)
	High (>221)	258 (38.91)	106,769 (36.93)	473 (38.64)	33,285 (27.64)	554 (40.56)	52,312 (31.12)
	Mean [g/day (SD)]	205.48 (137.14)	207.22 (146.95)	201.13 (135.44)	183.18 (146.82)	213.87 (153.47)	190.83 (142.55)
P [a]		0.556		<0.001		<0.001	
P [b]							

Table 5-1_continued *Characteristics of the Study Population (3,253 cases and 577,700 non-cases) and Vegetable Intake of Included Studies*

Characteristics		Ethnicity (%)		Total Energy [kcal/day (SD)]	Alcohol [g/day (SD)]	Sugar [g/day (SD)]	Meat [g/day (SD)]
		Caucasian	Non-Caucasian				
	Total	570,382	10,292	580,953	580,953	580,953	580,953
Vegetable Intake (g/day)	Low (0-117)	184,093 (32.28)	9,363 (90.97)	2,229.87 (1939.18)	4.28 (10.20)	23.20 (39.85)	70.63 (81.39)
	Medium (117-221)	192,822 (33.81)	764 (7.42)	2,071.28 (670.85)	3.19 (8.34)	23.55 (46.67)	76.63 (53.73)
	High (>221)	193,467 (33.92)	165 (1.60)	2,217.35 (713.94)	2.20 (6.43)	27.55 (64.98)	85.65 (61.73)
	Mean [g/day (SD)]	200.22 (145.76)	49.29 (53.49)	197.50 (146.00)	197.50 (146.00)	197.50 (146.00)	197.50 (146.00)
P [a]		<0.001	0.592	<0.001	<0.001	<0.001	<0.001
P [b]							

Table 5-1_continued *Characteristics of the Study Population (3,253 cases and 577,700 non-cases) and Vegetable Intake of Included Studies*

Characteristics		Grain [g/day (SD)]	Fruit [g/day (SD)]	Fat [g/day (SD)]
	Total	580,953	580,953	580,953
	Low (0-117)	289.53 (295.21)	80.38 (85.44)	25.45 (26.45)
	Medium (117-221)	294.85 (177.10)	118.59 (102.57)	23.64 (16.10)
Vegetable Intake (g/day)	High (>221)	296.94 (174.82)	168.48 (131.14)	28.37 (20.28)
	Mean [g/day (SD)]	197.50 (146.00)	197.50 (146.00)	197.50 (146.00)
P [a]		0.003	0.286	<0.001
P [b]				

Table 5-1_continued *Characteristics of the Study Population (3,253 cases and 577,700 non-cases) and Vegetable Intake of Included Studies*

Characteristics		N (%)			Characteristics		N (%)		
			Case (%)	Non-case (%)				Case (%)	Non-case (%)
	Total	580,953	3,253	577,700		Total	97,622	1,514	96,108
Non-starchy Vegetable Intake (g/day)	Low (0-117)	193,651 (33.33)	1,012 (31.11)	192,639 (33.35)	**Starchy Vegetable Intake (g/day)**	Low (0-117)	32,698 (33.33)	692 (45.71)	31,849 (33.14)
	Medium (117-221)	193,651 (33.33)	984 (30.25)	192,667 (33.35)		Medium (117-221)	32,698 (33.49)	491 (32.43)	32,207 (33.51)
	High (>221)	193,651 (33.33)	1,257 (38.64)	192,394 (33.30)		High (>221)	32,383 (33.17)	331 (21.86)	32,052 (33.35)
	Mean [g/day (SD)]	194.52 (146.02)	201.44 (141.33)	194.49 (146.04)		Mean [g/day (SD)]	17.69 (20.15)	12.74 (12.74)	17.77 (20.20)
P [a]			<0.001		**P [a]**			<0.001	
P [b]					**P [b]**				

Table 5-1_continued *Characteristics of the Study Population (3,253 cases and 577,700 non-cases) and Vegetable Intake of Included Studies*

Characteristics		N (%)			Characteristics		N (%)		
			Case (%)	Non-case (%)				Case (%)	Non-case (%)
	Total	574,381	3,212	96,108		Total	574,381	3,212	96,108
Green Leafy Vegetable Intake	Low (0-117)	199,921 (34.81)	952 (29.64)	198,969 (34.84)	**Cruciferous Vegetable Intake**	Low (0-117)	191,463 (33.33)	895 (27.86)	190,568 (33.36)
	Medium (117-221)	183,007 (31.86)	1,037 (32.29)	181,970 (31.86)		Medium (117-221)	191,459 (33.33)	878 (27.33)	190,581 (33.37)
	High (>221)	191,453 (33.33)	1,223 (38.08)	190,230 (33.31)		High (>221)	191,459 (33.33)	1,439 (44.80)	190,020 (33.27)
	Mean [g/day (SD)]	23.85 (35.29)	24.26 (36.29)	23.85 (35.28)		Mean [g/day (SD)]	26.69 (36.28)	33.95 (42.46)	26.65 (36.24)
P [a]			0.519		**P** [a]			<0.001	
P [b]					**P** [b]				

[a] *Calculated by Chi-square (χ^2) test for categorical variables and t-test for continuous variables between bladder cancer cases and non-cases. Age at the time of recruitment.*

Abbreviation: N=Number; SD, Standard Deviation; kcal, kilocalorie; g, gram.

P^a: *$P\text{-}_{value}$ <0.05 was considered statistically significant.*

P^b: *$P\text{-}_{interaction}$ <0.05 was considered statistically significant.*

Table 5-2 *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Intake of Vegetable Stratified by Subtype, Sex and Smoking Status*

Study Subgroup		**Overall**			**NMIBC**			**MIBC**		
Model Adjustments		No. Cases	Model 1 [d]	Model 2 [d, e]	No. Cases	Model 1 [d]	Model 2 [d, e]	No. Cases	Model 1 [d]	Model 2 [d, e]
Vegetable Intake (g/day)	Low (0-117)	997	Ref.	Ref.	392	Ref.	Ref.	167	Ref.	Ref.
	Medium (117-221)	971	1.00 (0.91-1.10)	1.00 (0.91-1.10)	390	1.10 (0.94-1.28)	1.10 (0.94-1.28)	213	1.12 (0.89-1.41)	1.12 (0.89-1.41)
	High (>221)	1,285	0.96 (0.86-1.07)	0.95 (0.85-1.07)	485	1.02 (0.85-1.22)	1.01 (0.84-1.21)	393	1.00 (0.78-1.28)	0.99 (0.77-1.27)
HR Per 1 SD Increase (95% CI)			0.98 (0.94-1.03)	0.97 (0.93-1.02)		0.98 (0.90-1.06)	0.97 (0.89-1.05)		0.98 (0.89-1.08)	0.97 (0.88-1.07)
P-trend			0.424	0.296		0.586	0.404		0.682	0.582

Table 5-2_continued *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Intake of Vegetable Stratified by Subtype, Sex and Smoking Status*

Study Subgroup		**Male**			**Female**			**Never Smokers**		
Model Adjustments		No. Cases	Model 1 [d]	Model 2 [d, e]	No. Cases	Model 1 [d]	Model 2 [d, e]	No. Cases	Model 1 [d]	Model 2 [d, e]
	Low (0-117)	742	Ref.	Ref.	255	Ref.	Ref.	196	Ref.	Ref.
Vegetable Intake (g/day)	Medium (117-221)	688	1.05 (0.93-1.17)	1.02 (0.90-1.17)	283	0.89 (0.74-1.06)	0.89 (0.74-1.07)	209	1.03 (0.83-1.27)	1.03 (0.83-1.27)
	High (>221)	1,015	1.03 (0.90-1.18)	1.02 (0.90-1.17)	270	0.79 (0.64-0.98)	0.79 (0.64-0.98)	258	1.07 (0.84-1.36)	1.07 (0.84-1.36)
HR Per 1 SD Increase (95% CI)			1.01 (0.95-1.06)	1.00 (0.94-1.05)		0.89 (0.81-0.99)	0.89 (0.81-0.99)		1.02 (0.92-1.12)	1.01 (0.92-1.10)
P-trend			0.834	0.911		0.034	0.037		0.725	0.795

Table 5-2_continued *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Intake of Vegetable Stratified by Subtype, Sex and Smoking Status*

Study Subgroup		**Current Smoker**			**Former Smoker**		
Model Adjustments		No. Cases	Model 1 [d]	Model 2 [d, e]	No. Cases	Model 1 [d]	Model 2 [d, e]
Vegetable Intake (g/day)	Low (0-117)	391	Ref.	Ref.	410	Ref.	Ref.
	Medium (117-221)	360	1.00 (0.85-1.16)	0.99 (0.85-1.16)	402	1.00 (0.86-1.16)	1.00 (0.86-1.17)
	High (>221)	473	0.95 (0.79-1.14)	0.94 (0.78-1.13)	554	0.92 (0.78-1.10)	0.92 (0.77-1.10)
HR Per 1 SD Increase (95% CI)			0.92 (0.85-1.00)	0.91 (0.84-0.99)		0.99 (0.95-1.09)	1.00 (0.94-1.11)
P-trend			0.051	0.028		0.612	0.683

[b] Model 1: Adjusted for age (years, continuous), sex (male or female, if applicable), smoking (if applicable, smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), and total energy intake (kcal/day, continuous).
[b, c] Model 2: Additionally, ethnicity (Caucasian or non-Caucasian, if applicable) , alcohol intake (ml/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), and grain intake (g/day, continuous).
Abbreviation: NMIBC, Non-Muscle Invasive Bladder Cancer; MIBC, Muscle Invasive Bladder Cancer; SD, standard deviation; kcal, kilocalorie; g, gram.
Reference group was lowest intake (tertile 1).
P-trend <0.05 was considered statistically significant.

Subgroups of vegetable intake and bladder cancer risk

No evidence of association was found between the intake of non-starchy ($HR_{model2;\ highest\ vs\ lowest\ tertile}$=0.95, 95% CI=0.85-1.06, P=0.348 for trend, HR per 1 SD increment=0.98, 95% CI=0.93-1.03), starchy ($HR_{model2;\ highest\ vs\ lowest\ tertile}$=0.95, 95% CI=0.82-1.09, P=0.209 for trend, HR per 1 SD increment=0.96, 95% CI=0.89-1.02), green leafy or cruciferous vegetable ($HR_{model2;\ highest\ vs\ lowest\ tertile}$=0.95, 95% CI=0.84-1.07, P=0.866 for trend, HR per 1 SD increment=1.00, 95% CI=0.95-1.03), and the risk of the bladder cancer in the overall population (**Table 5-3** and **Table 5-4**). Both bladder cancer subtype and smoking status stratified analyses showed similar results that no association for any of the vegetables subtype intake and bladder cancer risk could be observed. Among females, however, an inverse association with bladder cancer risk was observed for the intake of non-starchy vegetable ($HR_{model2;\ highest\ vs\ lowest\ tertile}$=0.79, 95% CI=0.64-0.98, P=0.037 for trend; HR per 1 SD increment=0.89, 95% CI=0.81-0.99). No other vegetable subtype intake showed a significant association with bladder cancer risk neither in females nor in males.

Individual vegetable type intake and bladder cancer risk

A significant decreased bladder cancer risk was observed for high intake of curly kale ($HR_{model2;\ highest\ vs\ lowest\ tertile}$=0.80, 95% CI=0.68-0.95; $HR_{model2;\ medium\ vs\ lowest\ tertile}$=0.86, 95% CI=0.73-1.00, P=0.072 for trend; HR per 1 SD increment=0.94, 95% CI=0.87-1.01). No other individual

vegetable type showed a significant association with bladder cancer risk (**Supplementary Table 5-2**).

Vegetable diversity intake and bladder cancer risk

The HRs and 95% CIs for different levels for the DDS and bladder cancer risk are presented in **Table 5-5**. There was no evidence of significant associations for DDS with BC risk (consuming 6-8 individual types versus 0-2 individual types: HR_{model2}=0.95, 95% CI=0.62-1.45, P=0.342 for trend).

Sensitivity Analysis

Similar results for the intake of total vegetable with subgroups and bladder cancer risk was obtained from the sensitivity analysis when excluding bladder cancers diagnosed within the first 2 years after study entry (**Supplementary Table 5-3**). In addition, only the EPIC study showed a reduced risk of bladder cancer with higher intake of vegetable ($HR_{model2;\ highest\ vs\ lowest\ tertile}$=0.69, 95% CI=0.61-0.79, P<0.001 for trend), while other participating cohort studies showed null associations. The meta-analysis presented similar null association as our main analysis pooling original data from multiple studies ($HR_{model2;\ highest\ vs\ lowest\ tertile}$=0.91, 95% CI=0.71-1.12).

Table 5-3 *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Intake of Non-starchy or Starchy Vegetable*

Subgroup		Overall			NMIBC			MIBC		
Models		No. Cases	Model 1 [d]	Model 2 [d, e]	No. Cases	Model 1 [d]	Model 2 [d, e]	No. Cases	Model 1 [d]	Model 2 [d, e]
Non-starchy Vegetable Intake (g/day)	Low (0-108)	1,012	Ref.	Ref.	407	Ref.	Ref.	175	Ref.	Ref.
	Medium (108-211)	984	0.99 (0.90-1.09)	0.99 (0.90-1.09)	391	1.05 (0.93-1.20)	1.05 (0.92-1.20)	223	1.03 (0.80-1.31)	1.02 (0.80-1.31)
	High (>211)	1,257	0.96 (0.86-1.07)	0.95 (0.85-1.06)	469	0.97 (0.80-1.17)	0.96 (0.79-1.16)	375	0.99 (0.84-1.16)	0.99 (0.84-1.16)
HR Per 1 SD Increase (95% CI)			0.98 (0.94-1.03)	0.98 (0.93-1.03)		0.99 (0.91-1.07)	0.97 (0.89-1.06)		0.98 (0.89-1.09)	0.97 (0.88-1.08)
P-trend			0.495	0.348		0.722	0.404		0.713	0.604
Starchy Vegetable Intake (g/day)	Low	692	Ref.	Ref.	357	Ref.	Ref.	290	Ref.	Ref.
	Medium (7-18)	491	0.94 (0.83-1.05)	0.94 (0.84-1.06)	235	0.83 (0.70-0.98)	0.83 (0.71-0.99)	226	1.05 (0.88-1.25)	1.05 (0.88-1.25)
	High (>18)	331	0.94 (0.81-1.08)	0.95 (0.82-1.09)	194	0.94 (0.77-1.13)	0.95 (0.78-1.15)	124	0.88 (0.70-1.11)	0.89 (0.70-1.12)
HR Per 1 SD Increase (95% CI)			0.95 (0.89-	0.96 (0.89-		0.93 (0.85-	0.93 (0.85-		0.97 (0.87-	0.98 (0.88-
P-trend			0.153	0.209		0.107	0.131		0.561	0.639

Table 5-3_continued *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Intake of Non-starchy or Starchy Vegetable*

Subgroup		Male			Female			Never Smokers		
Models		No. Cases	Model 1 [d]	Model 2 [d, e]	No. Cases	Model 1 [d]	Model 2 [d, e]	No. Cases	Model 1 [d]	Model 2 [d, e]
Non-starchy Vegetable Intake (g/day)	Low (0-108)	756	Ref.	Ref.	256	Ref.	Ref.	198	Ref.	Ref.
	Medium (108-211)	691	1.02 (0.91-1.15)	1.02 (0.91-1.15)	293	0.89 (0.75-1.07)	0.90 (0.75-1.08)	220	1.07 (0.87-1.32)	1.07 (0.87-1.32)
	High (>211)	998	1.03 (0.90-1.17)	1.02 (0.89-1.16)	259	0.79 (0.64-0.98)	0.79 (0.64-0.98)	245	1.09 (0.85-1.38)	1.08 (0.85-1.38)
HR Per 1 SD Increase (95% CI)			1.01 (0.95-1.07)	1.00 (0.94-1.06)		0.89 (0.80-0.99)	0.89 (0.81-0.99)		1.02 (0.93-1.13)	1.01 (0.92-1.10)
P-trend			0.758	0.981		0.039	0.037		0.628	0.795
Starchy Vegetable Intake (g/day)	Low (0-7)	571	Ref.	Ref.	121	Ref.	Ref.	112	Ref.	Ref.
	Medium (7-18)	406	0.96 (0.84-1.09)	0.96 (0.84-1.09)	85	0.85 (0.64-1.12)	0.85 (0.64-1.12)	91	0.87 (0.66-1.15)	0.87 (0.66-1.16)
	High (>18)	245	0.93 (0.79-1.09)	0.94 (0.80-1.11)	86	0.95 (0.70-1.30)	0.96 (0.71-1.31)	91	0.86 (0.64-1.17)	0.86 (0.64-1.17)
HR Per 1 SD Increase (95% CI)			0.96 (0.89-1.04)	0.97 (0.89-1.04)		0.93 (0.80-1.07)	0.93 (0.81-1.07)		0.93 (0.82-1.07)	0.90 (0.79-1.03)
P-trend			0.307	0.392		0.288	0.310		0.310	0.116

Table 5-3_continued *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Intake of Non-starchy or Starchy Vegetable*

Subgroup		Current Smokers			Former Smokers		
Models		No. Cases	Model 1 [d]	Model 2 [d, e]	No. Cases	Model 1 [d]	Model 2 [d, e]
Non-starchy Vegetable Intake (g/day)	Low (0-108)	385	Ref.	Ref.	429	Ref.	Ref.
	Medium (108-211)	368	1.01 (0.87-1.18)	1.01 (0.86-1.18)	396	0.93 (0.80-1.09)	0.94 (0.80-1.09)
	High (>211)	471	0.97 (0.80-1.16)	0.95 (0.79-1.14)	541	0.90 (0.75-1.07)	0.90 (0.75-1.07)
HR Per 1 SD Increase (95% CI)			0.92 (0.85-1.00)	0.91 (0.84-0.99)		0.99 (0.95-1.09)	1.00 (0.94-1.11)
P-trend			0.051	0.028		0.612	0.683
Starchy Vegetable Intake (g/day)	Low (0-7)	270	Ref.	Ref.	310	Ref.	Ref.
	Medium (7-18)	149	0.93 (0.76-1.14)	0.93 (0.76-1.14)	251	0.95 (0.81-1.13)	0.96 (0.81-1.13)
	High (>18)	67	0.98 (0.74-1.29)	0.98 (0.74-1.30)	173	0.94 (0.77-1.15)	0.95 (0.78-1.16)
HR Per 1 SD Increase (95% CI)			0.97 (0.84-1.12)	0.92 (0.79-1.07)		0.95 (0.87-1.04)	0.96 (0.87-1.05)
P-trend			0.691	0.299		0.291	0.357

[b] *Model 1: Adjusted for age (years, continuous), sex (male or female, if applicable), smoking (if applicable, smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), and total energy intake (kcal/day, continuous).*

[b, c] *Model 2: Additionally, ethnicity (Caucasian or non-Caucasian, if applicable) , alcohol intake (ml/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), and grain intake (g/day, continuous).*

Abbreviation: SD, standard deviation; m, millilitre; kcal, kilocalorie; g, gram.

Reference group was lowest intake (tertile 1).

P-trend <0.05 was considered statistically significant.

Table 5-4 *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Intake of Green Leafy Vegetable or Cruciferous Vegetable*

Subgroup		Overall			NMIBC			MIBC		
Models		No. Cases	Model 1 [d]	Model 2 [e, f]	No. Cases	Model 1 [d]	Model 2 [e, f]	No. Cases	Model 1 [d]	Model 2 [e, f]
Green Leafy Vegetable Intake (g/day)	Low (0-5)	1,223	Ref.	Ref.	406	Ref.	Ref.	165	Ref.	Ref.
	Medium (5-23)	1,037	0.99 (0.90-1.09)	0.99 (0.90-1.09)	438	0.99 (0.85-1.15)	0.99 (0.85-1.15)	240	0.92 (0.77-1.10)	0.93 (0.77-1.11)
	High (>23)	952	0.95 (0.84-1.07)	0.95 (0.84-1.07)	423	0.95 (0.78-1.15)	0.95 (0.78-1.16)	368	0.82 (0.62-1.08)	0.99 (0.61-1.07)
HR Per 1 SD Increase (95% CI)			1.00 (0.96-1.06)	1.00 (0.95-1.03)		0.98 (0.93-1.03)	0.97 (0.89-1.06)		0.98 (0.86-1.12)	0.98 (0.86-1.12)
***P*-trend**			0.843	0.866		0.465	0.404		0.771	0.744
Cruciferous Vegetable Intake (g/day)	Low (0-7)	895	Ref.	Ref.	336	Ref.	Ref.	165	Ref.	Ref.
	Medium (7-18)	878	0.96 (0.87-	0.96 (0.87-	382	0.92 (0.78-	0.91 (0.78-	240	0.94 (0.71-	0.95 (0.71-
	High (>18)	1,439	0.93 (0.82-1.05)	0.93 (0.82-1.05)	549	0.88 (0.72-1.07)	0.95 (0.78-1.15)	368	0.92 (0.71-1.19)	0.93 (0.72-1.20)
HR Per 1 SD Increase (95% CI)			0.99 (0.95-1.03)	0.99 (0.95-1.02)		0.98 (0.90-1.06)	0.93 (0.85-1.02)		0.97 (0.89-1.06)	0.97 (0.89-1.06)
P-trend			0.598	0.570		0.548	0.131		0.506	0.479

Table 5-4_continued *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Intake of Green Leafy Vegetable or Cruciferous Vegetable*

Subgroup		Male			Female			Never Smokers		
Models		No. Cases	Model 1 [d]	Model 2 [e, f]	No. Cases	Model 1 [d]	Model 2 [e, f]	No. Cases	Model 1 [d]	Model 2 [e, f]
Green Leafy Vegetable Intake (g/day)	Low (0-5)	724	Ref.	Ref.	228	Ref.	Ref.	198	Ref.	Ref.
	Medium (5-23)	763	0.99 (0.89-1.11)	1.00 (0.89-1.11)	274	1.00 (0.79-1.26)	0.99 (0.78-1.25)	215	1.00 (0.81-1.24)	1.00 (0.81-1.24)
	High (>23)	927	0.93 (0.81-1.07)	0.93 (0.81-1.07)	296	1.00 (0.83-1.20)	0.99 (0.82-1.19)	243	0.87 (0.67-1.13)	0.88 (0.67-1.14)
HR Per 1 SD Increase (95% CI)			0.99 (0.95-1.07)	0.98 (0.95-1.07)		0.90 (0.81-1.01)	0.91 (0.82-1.01)		0.95 (0.85-1.06)	0.95 (0.85-1.06)
P-trend			0.141	0.159		0.064	0.072		0.389	0.384
Cruciferous Vegetable Intake (g/day)	Low (0-7)	724	Ref.	Ref.	228	Ref.	Ref.	167	Ref.	Ref.
	Medium (7-18)	763	0.98 (0.87-1.11)	0.98 (0.87-1.11)	274	0.93 (0.74-1.17)	0.93 (0.74-1.18)	198	1.08 (0.86-1.35)	0.87 (0.66-1.16)
	High (>18)	927	0.93 (0.80-	0.93 (0.80-	296	0.90 (0.74-	0.90 (0.74-	291	1.01 (0.78-	0.86 (0.64-
HR Per 1 SD Increase (95% CI)			0.95 (0.88-1.10)	0.95 (0.89-1.10)		0.90 (0.81-1.01)	0.91 (0.83-1.00)		0.94 (0.85-1.03)	0.90 (0.79-1.03)
P-trend			0.588	0.648		0.064	0.060		0.187	0.116

Table 5-4_continued *Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Intake of Green Leafy Vegetable or Cruciferous Vegetable*

Subgroup		Current Smokers			Former Smokers		
Models		No. Cases	Model 1 [d]	Model 2 [e, f]	No. Cases	Model 1 [d]	Model 2 [e, f]
Green Leafy Vegetable Intake (g/day)	Low (0-5)	328	Ref.	Ref.	426	Ref.	Ref.
	Medium (5-23)	439	1.07 (0.92-1.24)	1.07 (0.92-1.24)	383	0.98 (0.81-1.18)	0.97 (0.81-1.17)
	High (>23)	430	0.96 (0.80-1.16)	0.97 (0.80-1.17)	550	0.92 (0.79-1.07)	0.92 (0.79-1.07)
HR Per 1 SD Increase (95% CI)			0.96 (0.86-1.02)	0.95 (0.86-1.05)		0.99 (0.94-1.10)	1.00 (0.94-1.13)
P-trend			0.372	0.347		0.110	0.113
Cruciferous Vegetable Intake (g/day)	Low (0-7)	365	Ref.	Ref.	363	Ref.	Ref.
	Medium (7-18)	344	1.04 (0.88-1.22)	1.04 (0.88-1.22)	336	0.93 (0.77-1.11)	0.95 (0.80-1.12)
	High (>18)	488	0.89 (0.73-1.09)	0.89 (0.72-1.09)	660	0.85 (0.72-1.00)	0.89 (0.78-1.01)
HR Per 1 SD Increase (95% CI)			0.97 (0.90-1.16)	0.97 (0.90-1.05)		0.97 (0.88-1.11)	0.96 (0.87-1.10)
P-trend			0.528	0.499		0.496	0.520

[b] Model 1: Adjusted for age (years, continuous), sex (male or female, if applicable), smoking (if applicable, smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), and total energy intake (kcal/day, continuous).
[b, c] Model 2: Additionally, ethnicity (Caucasian or non-Caucasian, if applicable) , alcohol intake (ml/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), and grain intake (g/day, continuous).
Abbreviation: SD, standard deviation; ml, millilitre; kcal, kilocalorie; g, gram.
Reference group was lowest intake (tertile 1).
P-trend <0.05 was considered statistically significant.

***Table 5-5** Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Diet Diversity Scores of Vegetable*

Models	Diet Diversity Scores of Vegetable				P-trend
	0-2	2-4	4-6	6-8	
No. Cases	27	964	1,944	277	
Model 1 [f]	Ref.	0.87 (0.59-1.29)	0.91 (0.61-1.37)	0.94 (0.61-1.43)	0.371
Model 2 [f, g]	Ref.	0.87 (0.59-1.29)	0.92 (0.61-1.37)	0.95 (0.62-1.45)	0.342

[f] Model 1: Adjusted for age (years, continuous), sex (male or female, if applicable), smoking (if applicable, smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), and total energy intake (kcal/day, continuous), and total vegetable intake.
[f, g] Model 2: Additionally, ethnicity (Caucasian or non-Caucasian, if applicable) , alcohol intake (ml/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), grain intake (g/day, continuous).
Abbreviation: NMIBC, Non-Muscle Invasive Bladder Cancer; MIBC, Muscle Invasive Bladder Cancer; SD, standard deviation; kcal, kilocalorie; g, gram.
Reference group was lowest intake (tertile 1).
P-trends <0.05 was considered statistically significant.

Discussion

In this large multi-centre prospective cohort study, we found an inverse association between intake of total and non-starchy vegetable and bladder cancer risk among females. This association, however, was not observed among males or for the overall population.

The observed null associations in the overall population and in males are in line with two previously conducted meta-analyses, including cohort studies only [26, 29], also showing no or little evidence of a protective effect of vegetable intake on bladder cancer risk. However, two other meta-analyses, including both case-control studies and cohort studies, showed a protective effect of vegetables intake and bladder cancer risk [27, 28]. A possible explanation for this observed discrepancy in results might be the inclusion of different study designs. Results obtained from long-term cohort studies, who prospectively obtained data, are suggested of substantially superior quality to those obtained from case-control studies. Retrospectively obtaining dietary data in case-control studies may result is false findings, since cases usually recall their dietary intake better than controls [63]. Another explanation could be that "total vegetables intake" may vary in its constituents *(i.e.* different individual vegetable types included) across different studies. However, our DDS analyses showed no association between the vegetable diversity and bladder cancer risk.

In the present study, a reduced risk of bladder cancer was found among females consuming high levels of "total vegetable" and non-starchy vegetables. Similar results were shown by other large prospective cohort studies, including EPIC, which also showed a protective effect of total vegetables intake among females only [22, 64, 65]. In addition, results from a previous BLEND study investigating the influence of vegetables intake on bladder cancer in case-control studies showed an inverse association among females, while a borderline decreased risk of bladder cancer was observed among males.

The inverse association found among females but not among males may be partially explained by differences in hormonal factors (*e.g.* oestrogen) and different urination habits between males and females. Experimental studies showed that nutrients or chemicals obtained from vegetable have the potential to modulate the effect of endogenous hormones and thereby, inhibit the development of bladder cancer among females [66]. However, it cannot be

excluded that residual confounding by other factors might explain the sex difference. In addition, it should be note that the statistical power was rather limited for females compared to males (808 incident bladder cancers in females and 2,445 incident bladder cancers in males). Future research is, therefore, needed to clarify this finding and provide better understanding of how vegetable influence bladder cancer risk in terms of sexes.

The WCRF/AICR CUP reported a suggestive but not conclusive protective effects of non-starchy vegetable on bladder cancer risk [67, 68]. In addition, by summarizing 10 primary studies on non-starchy vegetable intake and bladder cancer risk, the WCRF/AICR CUP reported a borderline decreased bladder cancer risk per 1 serving/day increment of non-starchy vegetable intake (RR=0.97, 95% CI=0.94-1.00). The present study supports these findings by showing that a higher intake of non-starchy vegetable decreases the risk of bladder cancer among females. However, the biological mechanism for the role of non-starchy vegetable in the process of bladder carcinogenesis remains unclear.

In contrast to the protective effect of non-starchy vegetable on bladder cancer, a null association was observed for starchy vegetables. This finding is in line with a previous case-control study, which also reported a null association between higher starchy vegetable intake and bladder cancer risk. Starchy vegetables can raise blood sugar levels faster than non-starchy vegetables, resulting in an increased glycaemic loading and insulin response, and may thereby resist the suggested beneficial effects of vegetable [69].

Green leafy vegetables contain several antioxidant nutrients, minerals, dietary fibre, phenols, flavonoids and phytochemicals [70-72], which may reduce oxidative stress and DNA damage caused by free radicals, and affect pathways controlling cell proliferation and apoptosis [73]. However, surprisingly no protective effect of this vegetable subtype could be observed in the present study or in any previously conducted prospective cohort studies [74, 75]. Null association in epidemiological studies may be the result of a limited amount range of intake, and thereby insufficient power to detect an association.

Cruciferous vegetables, in particular broccoli, contain high levels of glucosinolates. During food preparation, glucosinolates are transformed into isothiocyanates, which are suggested

to favourably modify the carcinogen metabolism via inhibition of Phase 1 enzymes and/or induction of Phase 2 enzymes [76]. Several pooled results from case-control studies indeed showed a protective effect of cruciferous vegetables intake and bladder cancer risk [61, 77].

In the present study, however, this hypothesis could not be strengthened, since results showed a null association between cruciferous vegetable intake and bladder cancer risk. A similar null-finding was observed by the WCRF/AICR International's CUP [19], who conducted a meta-analysis on results from 7 primary cohort studies, and by a meta-meta-analysis conducted by *Al-Zalabani* et al. [61] who pooled results from 2 meta-analyses of cohort studies. The discrepancy, between these epidemiological studies, might again be caused by the different study design of the included primary studies. Further studies need to be conducted in order clarify the associations between subtypes of vegetable and bladder cancer risk and the potential mechanisms.

Although it is known that several individual vegetables types contain nutrients with a proven anti-carcinogenic effect, the present study only found a significant protective effect of high curly kale intake and bladder cancer risk. Curly kale, is reported to be one of the most nutrient-dense vegetable and might therefore help to prevent cancer [78, 79]. The exact mechanism of this protective effect on carcinogenesis, however, needs to be elucidated.

Null findings of other individual vegetable types might be caused by the fact that the intake of vegetable (sub)types individually does not reach a certain level which is needed to influence bladder cancer development. In addition, although BLEND provided enough statistical power for most individual vegetable subtypes, some subtypes (*i.e.* starchy vegetables) might have lacked sufficient power to detect small effect sizes. Future observational or experimental researches aiming to investigate the effect of the intake of individual vegetable subtypes in preventing bladder cancer are warranted.

The effect of vegetables on bladder cancer risk may interact with smoking. Several experimental studies have reported a group of natural phytochemicals uniquely present in vegetables can modulate the smoking-related carcinogen metabolism and facilitate carcinogen detoxification, thus inhibiting smoking-related carcinogenesis [80-83]. This could

potentially lead to a difference in the effect of vegetable intake on bladder cancer when stratifying for smoking status. However, in the present study the different smoking status categories, did not show different results. Further research is needed to elucidate the potential interactive effect of smoking with vegetable intake on bladder cancer risk.

For the present study, data were pooled from 14 prospective cohorts, to obtain a sample size, large enough to permit detailed analyses with good precision. The study, however, also has some limitations: a) other than age, sex and smoking, limited information was available on other possible risk factors, for the development of bladder cancer, such as *body mass index* (BMI), physical activity, socioeconomic status (SES), disinfection by-products, arsenic in the drinking water, and occupational exposures to potentially carcinogenic chemicals. Although, adjustments for these factors could have influenced the results, the current literature shows that only a small proportion of bladder cancer cases can be attributed to these factors [72]; b) some tumour subtype (*i.e.* muscle-invasive/non-muscle-invasive) information was missing, which hampered the statistical power required for stratified subgroup analyses; c) sample size for starchy vegetable was smaller than other vegetable subgroups, which may have caused insufficient statistical power for detailed analyses; d) although we found similar results after adjusting for potential dietary risk factors, it is still possible that the observed associations were confounded by other dietary constituents or additives associated with vegetable intake; e) for most cohorts, the exposure variable was assessed by FFQs. Therefore, measurement error and misclassification of study participants in terms of the exposure and outcome are unavoidable; f) the present study sample consists mostly of Caucasians, and this may limit the generalizability of our results to other racial/ethnic populations or geographic regions; g) there is limited mechanistic evidence to interpret our findings, therefore, further studies are needed to clarify whether there is an underlying mechanism for the association between intake of vegetable and bladder cancer risk, especially for females.

Conclusion

In summary, the present study, including over 3,200 cases of incident bladder cancer occurring in almost 0.6 million participants, indicates that a higher intake of total vegetable, particularly non-starchy vegetable and curly kale, may decrease the risk of bladder cancer

among females. Future studies are needed to clarify the exact mechanism of this protective effect and should consider sex differences when investigating the association between vegetable intake and bladder cancer risk.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69(1):7-34. doi:10.3322/caac.21551.

2. Richters A, Aben KKH, Kiemeney L. The global burden of urinary bladder cancer: an update. World J Urol. 2019. doi:10.1007/s00345-019-02984-4.

3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424. doi:10.3322/caac.21492.

4. Antoni S, Ferlay J, Soerjomataram I, Znaor A, Jemal A, Bray F. Bladder Cancer Incidence and Mortality: A Global Overview and Recent Trends. *Eur Urol* 2017;71(1):96-108. doi:10.1016/j.eururo.2016.06.010.

5. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011;61(2):69-90. doi:10.3322/caac.20107.

6. Ploeg M, Aben KK, Kiemeney LA. The present and future burden of urinary bladder cancer in the world. *World J Urol* 2009;27(3):289-293. doi:10.1007/s00345-009-0383-3.

7. Marugame T, Mizuno S. Comparison of prostate cancer mortality in five countries: France, Italy, Japan, UK and USA from the WHO mortality database (1960–2000). *Jpn J Clin Oncol* 2005;35(11):690-691. doi:10.1093/jjco/hyi185.

8. Stewart B, Wild CP. World cancer report 2014. *International Agency for Research on Cancer, World Health Organization*: WHO Press, 2017.

9. Letašiová S, Medveďová A, Šovčíková A *et al*. Bladder cancer, a review of the environmental risk factors. *Environmental Health* 2012;11(Suppl 1):S11. doi:10.1186/1476-069X-11-S1-S11.

10. Botteman MF, Pashos CL, Redaelli A, Laskin B, Hauser R. The health economics of bladder cancer. *Pharmacoeconomics* 2003;21(18):1315-1330. doi:10.1007/BF03262330.

11. Al-Zalabani AH, Stewart KF, Wesselius A, Schols AM, Zeegers MP. Modifiable risk factors for the prevention of bladder cancer: a systematic review of meta-analyses. *Eur J EPDEMIOL*. 2016;31(9):811-51.

12. Piyathilake C. Dietary factors associated with bladder cancer. Investigative and clinical urology. 2016;57 Suppl 1:S14-25. doi:10.4111/icu.2016.57.S1.S14.

13. Johansson SL, Cohen SM. Epidemiology and etiology of bladder cancer. Semin Surg Oncol. 1997;13(5):291-8. doi:10.1002/(sici)1098-2388(199709/10)13:5<291::aid-ssu2>3.0.co;2-8.

14. Thun M, Linet MS, Cerhan JR, Haiman CA, Schottenfeld D. Cancer epidemiology and prevention. Oxford University Press; 2017.

15. Oberoi S, Barchowsky A, Wu F. The global burden of disease for skin, lung, and bladder cancer caused by arsenic in food. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2014;23(7):1187-94. doi:10.1158/1055-9965.Epi-13-1317.

16. Kushi LH, Doyle C, McCullough M, Rock CL, Demark-Wahnefried W, Bandera EV et al. American Cancer Society Guidelines on nutrition and physical activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity. CA Cancer J Clin. 2012;62(1):30-67. doi:10.3322/caac.20140.

17. Bradbury KE, Appleby PN, Key TJ. Fruit, vegetable, and fiber intake in relation to cancer risk: findings from the European Prospective Investigation into Cancer and Nutrition (EPIC). The American journal of clinical nutrition. 2014;100 Suppl 1:394s-8s. doi:10.3945/ajcn.113.071357.

18. Aune D, Giovannucci E, Boffetta P, Fadnes LT, Keum N, Norat T et al. Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality-a systematic review and dose-response meta-analysis of prospective studies. Int J Epidemiol. 2017;46(3):1029-56. doi:10.1093/ije/dyw319.

19. World Cancer Research Fund International/American Institute for Cancer Research. Continuous Update Project Report: Diet, Nutrition, Physical Activity and Bladder Cancer. Available at: https://wwwwcrforg/int/continuous-update-project Accesed October 10, 2019.

20. 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. 1999;82(1):151-2. doi:10.1002/(sici)1097-0215(19990702)82:1<151::aid-ijc25>3.0.co;2-9.

21. Lin J, Kamat A, Gu J, Chen M, Dinney CP, Forman MR et al. Dietary intake of vegetables and fruits and the modification effects of GSTM1 and NAT2 genotypes on bladder cancer risk. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2009;18(7):2090-7. doi:10.1158/1055-9965.Epi-08-1174.

22. Park SY, Ollberding NJ, Woolcott CG, Wilkens LR, Henderson BE, Kolonel LN. Fruit and vegetable intakes are associated with lower risk of bladder cancer among women in the Multiethnic Cohort Study. J Nutr. 2013;143(8):1283-92. doi:10.3945/jn.113.174920.

23. Negri E, La Vecchia C, Franceschi S, D'Avanzo B, Parazzini F. Vegetable and fruit consumption and cancer risk. International journal of cancer. 1991;48(3):350-4. doi:10.1002/ijc.2910480307.

24. Peluso M, Airoldi L, Magagnotti C, Fiorini L, Munnia A, Hautefeuille A et al. White blood cell DNA adducts and fruit and vegetable consumption in bladder cancer. Carcinogenesis. 2000;21(2):183-7. doi:10.1093/carcin/21.2.183.

25. Chyou PH, Nomura AM, Stemmermann GN. A prospective study of diet, smoking, and lower urinary tract cancer. Annals of epidemiology. 1993;3(3):211-6. doi:10.1016/1047-2797(93)90021-u.

26. Vieira AR, Vingeliene S, Chan DS, Aune D, Abar L, Navarro Rosenblatt D et al. Fruits, vegetables, and bladder cancer risk: a systematic review and meta-analysis. Cancer Med. 2015;4(1):136-46. doi:10.1002/cam4.327.

27. Liu H, Wang XC, Hu GH, Guo ZF, Lai P, Xu L et al. Fruit and vegetable consumption and risk of bladder cancer: an updated meta-analysis of observational studies. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP). 2015;24(6):508-16. doi:10.1097/cej.0000000000000119.

28. Yao B, Yan Y, Ye X, Fang H, Xu H, Liu Y et al. Intake of fruit and vegetables and risk of bladder cancer: a dose-response meta-analysis of observational studies. Cancer Causes Control. 2014;25(12):1645-58. doi:10.1007/s10552-014-0469-0.

29. Xu C, Zeng XT, Liu TZ, Zhang C, Yang ZH, Li S et al. Fruits and vegetables intake and risk of bladder cancer: a PRISMA-compliant systematic review and dose-response meta-analysis of prospective cohort studies. Medicine. 2015;94(17):e759. doi:10.1097/md.0000000000000759.

30. Wang F, Shan Y. Sulforaphane retards the growth of UM-UC-3 xenographs, induces apoptosis, and reduces survivin in athymic mice. Nutr Res. 2012;32(5):374-80. doi:10.1016/j.nutres.2012.03.014.

31. Geng F, Tang L, Li Y, Yang L, Choi KS, Kazim AL et al. Allyl isothiocyanate arrests cancer cells in mitosis, and mitotic arrest in turn leads to apoptosis via Bcl-2 protein phosphorylation. The Journal of biological chemistry. 2011;286(37):32259-67. doi:10.1074/jbc.M111.278127.

32. Abbaoui B, Riedl KM, Ralston RA, Thomas-Ahner JM, Schwartz SJ, Clinton SK et al. Inhibition of bladder cancer by broccoli isothiocyanates sulforaphane and erucin:

characterization, metabolism, and interconversion. Molecular nutrition & food research. 2012;56(11):1675-87. doi:10.1002/mnfr.201200276.

33. Hu H, Zhang XP, Wang YL, Chua CW, Luk SU, Wong YC et al. Identification of a novel function of Id-1 in mediating the anticancer responses of SAMC, a water-soluble garlic derivative, in human bladder cancer cells. Molecular medicine reports. 2011;4(1):9-16. doi:10.3892/mmr.2010.380.

34. Wang YB, Qin J, Zheng XY, Bai Y, Yang K, Xie LP. Diallyl trisulfide induces Bcl-2 and caspase-3-dependent apoptosis via downregulation of Akt phosphorylation in human T24 bladder cancer cells. Phytomedicine : international journal of phytotherapy and phytopharmacology. 2010;17(5):363-8. doi:10.1016/j.phymed.2009.07.019.

35. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol. 1997;26 Suppl 1:S6-14. doi:10.1093/ije/26.suppl_1.s6.

36. Tjonneland A, Olsen A, Boll K, Stripp C, Christensen J, Engholm G et al. Study design, exposure variables, and socioeconomic determinants of participation in Diet, Cancer and Health: a population-based prospective cohort study of 57,053 men and women in Denmark. Scandinavian journal of public health. 2007;35(4):432-41. doi:10.1080/14034940601047986.

37. Clavel-Chapelon F, van Liere MJ, Giubout C, Niravong MY, Goulard H, Le Corre C et al. E3N, a French cohort study on cancer risk factors. E3N Group. Etude Epidemiologique aupres de femmes de l'Education Nationale. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP). 1997;6(5):473-8. doi:10.1097/00008469-199710000-00007.

38. Boeing H, Korfmann A, Bergmann MM. Recruitment procedures of EPIC-Germany. European Investigation into Cancer and Nutrition. Annals of nutrition & metabolism. 1999;43(4):205-15. doi:10.1159/000012787.

39. 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. doi:10.1079/phn2002394.

40. Panico S, Dello Iacovo R, Celentano E, Galasso R, Muti P, Salvatore M et al. Progetto ATENA, a study on the etiology of major chronic diseases in women: design, rationale and objectives. European journal of epidemiology. 1992;8(4):601-8. doi:10.1007/bf00146383.

41. Manjer J, Carlsson S, Elmstahl S, Gullberg B, Janzon L, Lindstrom M et al. The Malmo Diet and Cancer Study: representativity, cancer incidence and mortality in participants and non-participants. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP). 2001;10(6):489-99. doi:10.1097/00008469-200112000-00003.

42. 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. doi:10.1080/14034950310001432.

43. Beulens JW, Monninkhof EM, Verschuren WM, van der Schouw YT, Smit J, Ocke MC et al. Cohort profile: the EPIC-NL study. Int J Epidemiol. 2010;39(5):1170-8. doi:10.1093/ije/dyp217.

44. Davey GK, Spencer EA, Appleby PN, Allen NE, Knox KH, Key TJ. EPIC-Oxford: lifestyle characteristics and nutrient intakes in a cohort of 33 883 meat-eaters and 31 546 non meat-eaters in the UK. Public health nutrition. 2003;6(3):259-69. doi:10.1079/phn2002430.

45. Day N, Oakes S, Luben R, Khaw KT, Bingham S, Welch A et al. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. British journal of cancer. 1999;80 Suppl 1:95-103. doi:PMID: 10466767.

46. Lund E, Dumeaux V, Braaten T, Hjartaker A, Engeset D, Skeie G et al. Cohort profile: The Norwegian Women and Cancer Study--NOWAC--Kvinner og kreft. Int J Epidemiol. 2008;37(1):36-41. doi:10.1093/ije/dym137.

47. van den Brandt PA, Goldbohm RA, van 't Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. Journal of clinical epidemiology. 1990;43(3):285-95. doi:10.1016/0895-4356(90)90009-E.

48. White E, Patterson RE, Kristal AR, Thornquist M, King I, Shattuck AL et al. VITamins And Lifestyle cohort study: study design and characteristics of supplement users. Am J Epidemiol. 2004;159(1):83-93. doi:10.1093/aje/kwh010.

49. Giles GG, English DR. The Melbourne Collaborative Cohort Study. IARC scientific publications. 2002;156:69-70. doi:PMID: 12484128.

50. Milne RL, Fletcher AS, MacInnis RJ, Hodge AM, Hopkins AH, Bassett JK et al. Cohort Profile: The Melbourne Collaborative Cohort Study (Health 2020). Int J Epidemiol. 2017;46(6):1757-i. doi:10.1093/ije/dyx085.

51. Ozasa K, Shimizu Y, Sakata R, Sugiyama H, Grant EJ, Soda M et al. Risk of cancer and non-cancer diseases in the atomic bomb survivors. Radiation protection dosimetry. 2011;146(1-3):272-5. doi:10.1093/rpd/ncr168.

52. Goossens ME, Isa F, Brinkman M *et al*. International pooled study on diet and bladder cancer: the bladder cancer, epidemiology and nutritional determinants (BLEND) study: design and baseline characteristics. *Archives of Public Health* 2016;74(1):30-39. doi:10.1186/s13690-016-0140-1.

53. Bassett JK, English DR, Fahey MT, Forbes AB, Gurrin LC, Simpson JA et al. Validity and calibration of the FFQ used in the Melbourne Collaborative Cohort Study. Public health nutrition. 2016;19(13):2357-68. doi:10.1017/s1368980016000690.

54. Hodge A, Patterson AJ, Brown WJ, Ireland P, Giles G. The Anti Cancer Council of Victoria FFQ: relative validity of nutrient intakes compared with weighed food records in young to middle-aged women in a study of iron supplementation. Aust N Z J Public Health. 2000;24(6):576-83. doi:10.1111/j.1467-842x.2000.tb00520.x.

55. Kaaks R, Riboli E. Validation and calibration of dietary intake measurements in the EPIC project: methodological considerations. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol. 1997;26 Suppl 1:S15-25. doi:10.1093/ije/26.suppl_1.s15.

56. Zeegers MP, Goldbohm RA, van den Brandt PA. Are retinol, vitamin C, vitamin E, folate and carotenoids intake associated with bladder cancer risk? Results from the Netherlands Cohort Study. British journal of cancer. 2001;85(7):977-83. doi:10.1054/bjoc.2001.1968.

57. Ferrari P, Slimani N, Ciampi A, Trichopoulou A, Naska A, Lauria C et al. Evaluation of under- and overreporting of energy intake in the 24-hour diet recalls in the European Prospective Investigation into Cancer and Nutrition (EPIC). Public health nutrition. 2002;5(6b):1329-45. doi:10.1079/phn2002409.

58. Poortvliet E, Klensin J, Kohlmeier L. Rationale document for the Eurocode 2 food coding system (version 91/2). *Eur J Clin Nutr* 1992;46(Suppl 5):S9-S24.

59. Giganti MJ, Luz PM, Caro-Vega Y, Cesar C, Padgett D, Koenig S et al. A Comparison of Seven Cox Regression-Based Models to Account for Heterogeneity Across Multiple HIV Treatment Cohorts in Latin America and the Caribbean. AIDS research and human retroviruses. 2015;31(5):496-503. doi:10.1089/aid.2014.0241.

60. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. The American journal of clinical nutrition. 1997;65(4 Suppl):1220S-8S; discussion 9S-31S. doi:10.1093/ajcn/65.4.1220S.

61. Al-Zalabani AH, Stewart KF, Wesselius A, Schols AM, Zeegers MP. Modifiable risk factors for the prevention of bladder cancer: a systematic review of meta-analyses. European journal of epidemiology. 2016;31(9):811-51. doi:10.1007/s10654-016-0138-6.

62. Bagnardi V, Zambon A, Quatto P, Corrao G. Flexible meta-regression functions for modeling aggregate dose-response data, with an application to alcohol and mortality. *Am J Epidemiol* 2004;159(11):1077-1086.

63. Tarasuk VS, Brooker AS. Interpreting epidemiologic studies of diet-disease relationships. J Nutr. 1997;127(9):1847-52. doi:10.1093/jn/127.9.1847.

64. Buchner FL, Bueno-de-Mesquita HB, Ros MM, Kampman E, Egevad L, Overvad K et al. Consumption of vegetables and fruit and the risk of bladder cancer in the European Prospective Investigation into Cancer and Nutrition. International journal of cancer. 2009;125(11):2643-51. doi:10.1002/ijc.24582.

65. Ros MM, Bueno-de-Mesquita HB, Kampman E, Buchner FL, Aben KK, Egevad L et al. Fruit and vegetable consumption and risk of aggressive and non-aggressive urothelial cell carcinomas in the European Prospective Investigation into Cancer and Nutrition. Eur J Cancer. 2012;48(17):3267-77. doi:10.1016/j.ejca.2012.05.026.

66. Ntzeros K, Stamatakos M, Stokidis S, Louka G. Bladder Cancer: The Hormonal Dependence Enigma and a New Hormonal Player. Indian J Surg. 2015;77(Suppl 3):1502-3. doi:10.1007/s12262-014-1133-3.

67. WCRF/AICR:Discover how diet, nutrition and physical activity affect cancer risk. Available at: https://wwwwcrforg/dietandcancer/interactive-cancer-risk-matrix#download_block. Accessed October 10, 2019.

68. Continuous Update Project. Available at: https://wwwwcrforg/int/continuous-update-project. Accessed Dec 25th, 2019.

69. Cust AE, Skilton MR, van Bakel MM, Halkjaer J, Olsen A, Agnoli C et al. Total dietary carbohydrate, sugar, starch and fibre intakes in the European Prospective Investigation into Cancer and Nutrition. Eur J Clin Nutr. 2009;63 Suppl 4:S37-60. doi:10.1038/ejcn.2009.74.

70. Tanaka T, Shnimizu M, Moriwaki H. Cancer chemoprevention by carotenoids. Molecules. 2012;17(3):3202-42. doi:10.3390/molecules17033202.

71. Padayatty SJ, Katz A, Wang Y, Eck P, Kwon O, Lee JH et al. Vitamin C as an antioxidant: evaluation of its role in disease prevention. Journal of the American College of Nutrition. 2003;22(1):18-35. doi:10.1080/07315724.2003.10719272.

72. Kang ZC, Tsai SJ, Lee H. Quercetin inhibits benzo[a]pyrene-induced DNA adducts in human Hep G2 cells by altering cytochrome P-450 1A1 gene expression. Nutr Cancer. 1999;35(2):175-9. doi:10.1207/s15327914nc352_12.

73. Liu RH. Dietary bioactive compounds and their health implications. J Food Sci. 2013;78 Suppl 1:A18-25. doi:10.1111/1750-3841.12101.

74. Michaud DS, Spiegelman D, Clinton SK, Rimm EB, Willett WC, Giovannucci EL. Fruit and vegetable intake and incidence of bladder cancer in a male prospective cohort. Journal of the National Cancer Institute. 1999;91(7):605-13. doi:10.1093/jnci/91.7.605.

75. Larsson SC, Andersson SO, Johansson JE, Wolk A. Fruit and vegetable consumption and risk of bladder cancer: a prospective cohort study. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2008;17(9):2519-22. doi:10.1158/1055-9965.Epi-08-0407.

76. Abbaoui B, Riedl KM, Ralston RA, Thomas‐Ahner JM, Schwartz SJ, Clinton SK et al. Inhibition of bladder cancer by broccoli isothiocyanates sulforaphane and erucin: characterization, metabolism, and interconversion. Molecular nutrition & food research. 2012;56(11):1675-87.

77. Liu B, Mao Q, Lin Y, Zhou F, Xie L. The association of cruciferous vegetables intake and risk of bladder cancer: a meta-analysis. World J Urol. 2013;31(1):127-33. doi:10.1007/s00345-012-0850-0.

78. Kuntz S, Kunz C. Extracts from Brassica oleracea L. convar. acephala var. sabellica inhibit TNF-alpha stimulated neutrophil adhesion in vitro under flow conditions. Food & function. 2014;5(6):1082-90. doi:10.1039/c3fo60562k.

79. Olsen H, Grimmer S, Aaby K, Saha S, Borge GI. Antiproliferative effects of fresh and thermal processed green and red cultivars of curly kale (Brassica oleracea L. convar. acephala var. sabellica). J Agric Food Chem. 2012;60(30):7375-83. doi:10.1021/jf300875f.

80. Tang L, Zirpoli GR, Jayaprakash V, Reid ME, McCann SE, Nwogu CE et al. Cruciferous vegetable intake is inversely associated with lung cancer risk among smokers: a case-control study. BMC cancer. 2010;10:162. doi:10.1186/1471-2407-10-162.

81. Hecht SS. Chemoprevention of cancer by isothiocyanates, modifiers of carcinogen metabolism. J Nutr. 1999;129(3):768s-74s. doi:10.1093/jn/129.3.768S.

82. Hecht SS. Inhibition of carcinogenesis by isothiocyanates. Drug Metab Rev. 2000;32(3-4):395-411. doi:10.1081/dmr-100102342.

83. Hecht SS, Carmella SG, Murphy SE. Effects of watercress consumption on urinary metabolites of nicotine in smokers. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 1999;8(10):907-13.

Supplementary Materials to Chapter 5

Supplementary Table 5-1 Additional Baseline Characteristic of the Participant Cohort Studies

Supplementary Table 5-2 Risk of Bladder Cancer According to Intake of Individual Types of Vegetable

Supplementary Table 5-3 Risk of Bladder Cancer According to Intake of Total and Subgroup Vegetable (After Removing 358 BC Cases Diagnosed within 2 years)

Supplementary Table 5-4 Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Intake of Vegetable Stratified by Study Centre

Supplementary Figure 5-1 Forest Plot of Meta-Analyses with HRs and 95% CIs for High vs. Low Intake of Total Vegetable with Bladder Cancer Risk

Supplementary Table 5-1 *Additional Baseline Characteristic of Included Studies*

Cohort	No. Participants (%) [a]	Follow-up Years [b]	Age at Baseline [c]	Person years	No. of Incidences	Years to Diagnosis [d]	% of Ever Smokers	Initiate year
EPIC (Europe)								
Denmark	55,999	10.93	50-66	610,894	391	6.82	0.65	1993
France	64,837	10.39	42-71	671,953	31	7.25	0.29	1993
Germany	49,456	9.88	19-70	487,336	207	6.23	0.54	1994
Greece	25,268	9.51	19-84	240,054	50	5.59	0.44	1993
Italy	45,201	11.23	24-78	507,116	186	6.89	0.55	1992
Spain	40,763	12.06	29-70	491,344	152	7.75	0.45	1992
Sweden	49,306	13.12	29-74	645,733	302	8.23	0.51	1991
Netherlands	37,080	11.81	20-70	437,563	107	7.36	0.62	1993
United Kingdom	75,014	11.14	20-98	834,165	248	7.05	0.44	1993
Norway	33,835	9.73	41-56	328,860	24	5.81	0.64	1998
NLCS (Netherlands)	5,245	14.06	55-70	73,591	877	9.34	0.68	1986
VITAL (USA)	69,290	6.75	50-76	464,730	346	3.87	0.52	2000
MCCS (Australia)	23,087	22.86	35-76	527,781	291	14.74	0.40	990
RERF-LSS (Japan)	6,572	28.04	19-91	184,160	41	22.10	0.46	1960
Overall	580,953	11.22	19-99	6,505,27	3,253	8.24	0.50	1960-2000

Abbreviations: EPIC, European Prospective Investigation into Cancer and Nutrition; NLCS, The Netherlands Cohort Study; VITAL, VITamins And Lifestyle study; MCCS, Melbourne Collaborative Cohort Study; RFRE-LSS, Radiation Effects Research Foundation-Life Span Study.

[a] As a result of exclusion criteria, the cohort sizes and number of cases included in the pooled analyses may differ from original study-specific publications.

[b] Median years of the time interval from two years after study entry to the date of the last follow-up (e.g. date of death, lost to follow-up, or study exit, whichever came first).

[c] Age ranges at baseline (minimum to maximum).

[d] Time-to-bladder cancer defined by median years of study entry to the date of diagnosis among bladder cancer cases.

Supplementary Table 5-2 *Risk of Bladder Cancer According to Intake of Individual Types of Vegetable*

Subgroup		Lettuce (g/day)			Spin(g/day)			Brassicas (g/day)		
Intake Tertiles		Low (0-6)	Medium (6-16)	High (>16)	Low (0-5)	Medium (5-12)	High (>12)	Low (0-7)	Medium (7-25)	High (>25)
No. Case		544	418	206	294	287	296	979	1,095	1,138
	HR (95% CI)	Ref.	1.03 (0.90-1.17)	1.11 (0.94-1.31)	Ref.	0.97 (0.82-1.14)	1.02 (0.87-1.20)	Ref.	0.93 (0.85-1.03)	0.89 (0.80-1.00)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	1.07 (0.98-1.16)			1.01 (0.95-1.08)			0.98 (0.93-1.03)		
	P-trend	0.138			0.678			0.376		
	HR (95% CI)	Ref.	1.04 (0.91-1.18)	1.13 (0.96-1.34)	Ref.	0.96 (0.81-1.13)	1.01 (0.86-1.19)	Ref.	0.93 (0.85-1.03)	0.89 (0.80-1.00)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	1.08 (0.99-1.18)			1.01 (0.95-1.08)			0.98 (0.93-1.03)		
	P-trend	0.075			0.700			0.393		

Supplementary Table 5-2_continued *Risk of Bladder Cancer According to Intake of Individual Types of Vegetable*

Subgroup		**Cauliflower (g/day)**			**Broccoli (g/day)**			**Chinese Cabbage (g/day)**		
Intake Tertiles		Low (0-2)	Medium (2-7)	High (>7)	Low (0-3)	Medium (3-9)	High (>17)	Low (0)	Medium (0-2)	High (>2)
No. Case		202	267	699	228	221	188	137	99	110
	HR (95% CI)	Ref.	1.05 (0.87-1.26)	0.93 (0.78-1.11)	Ref.	1.08 (0.89-1.30)	0.92 (0.75-1.12)	Ref.	1.03 (0.80-1.34)	0.91 (0.70-1.17)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	0.97 (0.92-1.02)			0.97 (0.89-1.05)			0.96 (0.86-1.08)		
	P-trend	0.237			0.435			0.504		
	HR (95% CI)	Ref.	1.05 (0.87-1.26)	0.93 (0.79-1.11)	Ref.	1.09 (0.90-1.31)	0.93 (0.76-1.14)	Ref.	1.04 (0.80-1.35)	0.92 (0.71-1.20)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	0.97 (0.92-1.02)			0.97 (0.90-1.06)			0.97 (0.87-1.09)		
	P-trend	0.259			0.535			0.647		

Supplementary Table 5-2_continued *Risk of Bladder Cancer According to Intake of Individual Types of Vegetable*

Subgroup		Cabbage (g/day)			Brussels Sprouts (g/day)			Curly Kale (g/day)		
Intake Tertiles		Low (0-7)	Medium (7-26)	High (>26)	Low (0-4)	Medium (4-10)	High (>10)	Low (0-2)	Medium (2-4)	High (>4)
No. Case		1,220	865	490	372	447	349	334	264	279
	HR (95% CI)	Ref.	0.98 (0.89-1.07)	1.03 (0.90-1.18)	Ref.	1.01 (0.88-1.16)	0.93 (0.81-1.08)	Ref.	0.81 (0.69-0.95)	0.86 (0.73-1.00)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	1.00 (0.94-1.06)			0.96 (0.88-1.04)			0.94 (0.87-1.01)		
	P-trend	0.946			0.309			0.070		
	HR (95% CI)	Ref.	0.98 (0.89-1.07)	1.03 (0.90-1.18)	Ref.	1.01 (0.88-1.16)	0.94 (0.81-1.09)	Ref.	0.80 (0.68-0.95)	0.86 (0.73-1.00)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	1.00 (0.94-1.06)			0.96 (0.89-1.04)			0.94 (0.87-1.01)		
	P-trend	0.972			0.345			0.072		

Supplementary Table 5-2_continued *Risk of Bladder Cancer According to Intake of Individual Types of Vegetable*

Subgroup		Rhubarb (g/day)			Celery (g/day)			Asparagus (g/day)		
Intake Tertiles		Low (0)	Medium (0-2)	High (>2)	Low (0-1)	Medium (1-4)	High (>4)	Low (0-0.5)	Medium (0.5-2)	High (>2)
No. Case		538	118	221	100	91	100	114	81	96
	HR (95% CI)	Ref.	0.99 (0.81-1.21)	0.96 (0.82-1.13)	Ref.	1.03 (0.77-1.37)	1.16 (0.88-1.53)	Ref.	0.91 (0.68-1.21)	1.08 (0.82-1.42)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	0.97 (0.90-1.04)			1.11 (1.00-1.23)			1.06 (0.96-1.17)		
	P-trend	0.434			0.048			0.254		
	HR (95% CI)	Ref.	0.99 (0.81-1.22)	0.97 (0.83-1.14)	Ref.	1.04 (0.78-1.38)	1.18 (0.89-1.57)	Ref.	0.91 (0.68-1.21)	1.10 (0.83-1.45)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	0.97 (0.91-1.05)			1.12 (1.01-1.24)			1.07 (0.97-1.18)		
	P-trend	0.457			0.032			0.198		

Supplementary Table 5-2_continued *Risk of Bladder Cancer According to Intake of Individual Types of Vegetable*

Subgroup		Garlic (g/day)			Onion (g/day)			Leek (g/day)		
Intake Tertiles		Low (0-0.5)	Medium (0.5-1)	High (>1)	Low (0-11)	Medium (11-22)	High (>22)	Low (0-2)	Medium (2-11)	High (>11)
No. Case		107	107	77	442	213	222	283	300	294
	HR (95% CI)	Ref.	1.22 (0.93-1.60)	0.98 (0.73-1.33)	Ref.	0.97 (0.86-1.10)	1.08 (1.06-1.09)	Ref.	1.14 (0.97-1.34)	1.12 (0.95-1.32)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	0.97 (0.86-1.10)			0.96 (0.90-1.02)			1.05 (0.99-1.11)		
	P-trend	0.611			0.211			0.120		
	HR (95% CI)	Ref.	1.21 (0.92-1.59)	0.98 (0.72-1.34)	Ref.	1.03 (0.87-1.21)	0.92 (0.78-1.08)	Ref.	1.14 (0.97-1.35)	1.13 (0.96-1.33)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	0.97 (0.86-1.10)			0.96 (0.89-1.03)			1.05 (0.99-1.12)		
	P-trend	0.642			0.212			0.093		

Supplementary Table 5-2_continued *Risk of Bladder Cancer According to Intake of Individual Types of Vegetable*

Subgroup		Tuber (g/day)			Sweet potato (g/day)			Carrot (g/day)		
Intake Tertiles		Low (0)	Medium (0-40)	High (>40)	Low (0-6)	Medium (6-16)	High (>16)	Low (0)	Medium (0-3)	High (>3)
No. Case		1,734	216	916	138	65	88	570	571	373
	HR (95% CI)	Ref.	0.88 (0.61-1.28)	0.88 (0.63-1.23)	Ref.	0.76 (0.57-1.03)	1.12 (0.85-1.47)	Ref.	0.93 (0.83-1.04)	0.96 (0.84-1.10)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	1.00 (0.98-1.01)			1.07 (0.97-1.18)			0.97 (0.90-1.04)		
	P-trend	0.691			0.193			0.389		
	HR (95% CI)	Ref.	0.88 (0.60-1.27)	0.87 (0.62-1.22)	Ref.	0.77 (0.57-1.03)	1.14 (0.86-1.50)	Ref.	0.94 (0.84-1.06)	0.98 (0.84-1.12)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	1.00 (0.98-1.01)			1.08 (0.97-1.19)			0.98 (0.91-1.06)		
	P-trend	0.539			0.160			0.602		

Supplementary Table 5-2_continued *Risk of Bladder Cancer According to Intake of Individual Types of Vegetable*

Subgroup		Beetroot (g/day)			Tomato (g/day)			Aubergine (g/day)		
Intake Tertiles		Low (0-0.5)	Medium (0.5-2.5)	High (>2.5)	Low (0-18)	Medium (18-50)	High (>50)	Low (0)	Medium (0-1.5)	High (>1.5)
No. Case		355	148	665	649	538	327	138	48	105
	HR (95% CI)	Ref.	0.82 (0.67-0.99)	0.97 (0.85-1.11)	Ref.	1.03 (0.92-1.16)	1.01 (0.88-1.17)	Ref.	0.98 (0.70-1.36)	1.32 (1.01-1.70)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	0.98 (0.94-1.02)			1.01 (0.94-1.07)			1.03 (0.94-1.13)		
	P-trend	0.322			0.852			0.528		
	HR (95% CI)	Ref.	0.82 (0.67-1.00)	0.98 (0.85-1.12)	Ref.	1.04 (0.92-1.17)	1.03 (0.90-1.19)	Ref.	0.98 (0.70-1.37)	1.34 (1.03-1.75)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	0.98 (0.94-1.02)			1.01 (0.95-1.08)			1.03 (0.94-1.13)		
	P-trend	0.358			0.651			0.479		

Supplementary Table 5-2_continued *Risk of Bladder Cancer According to Intake of Individual Types of Vegetable*

Subgroup		Sweet Pepper (g/day)			Cucumber (g/day)			Courgette (g/day)		
Intake Tertiles		Low (0)	Medium (0-3)	High (>3)	Low (0-3)	Medium (3-12)	High (>12)	Low (0-1)	Medium (1-5)	High (>5)
No. Case		430	214	233	93	107	91	106	92	93
	HR (95% CI)	Ref.	1.08 (0.92-1.28)	1.02 (0.86-1.19)	Ref.	1.45 (1.09-1.91)	1.02 (0.86-1.19)	Ref.	1.14 (0.86-1.51)	1.31 (0.99-1.75)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	1.00 (0.93-1.07)			1.06 (0.96-1.18)			1.10 (0.99-1.21)		
	P-trend	0.893			0.233			0.078		
	HR (95% CI)	Ref.	1.08 (0.92-1.27)	1.02 (0.87-1.20)	Ref.	1.46 (1.11-1.94)	1.41 (1.05-1.91)	Ref.	1.14 (0.86-1.52)	1.35 (1.01-1.79)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	1.00 (0.93-1.07)			1.08 (0.97-1.20)			1.10 (1.00-1.22)		
	P-trend	0.998			0.170			0.061		

Supplementary Table 5-2_continued *Risk of Bladder Cancer According to Intake of Individual Types of Vegetable*

Subgroup		Cucurbita Squash (g/day)			Pod and Seed Vegetable (g/day)			Pea (g/day)		
Intake Tertiles		Low (0-6)	Medium (6-19)	High (>19)	Low (0-0.2)	Medium (0.2-0.5)	High (>0.5)	Low (0-2)	Medium (2-7)	High (>7)
No. Case		124	84	83	845	985	1,382	104	119	123
	HR (95% CI)	Ref.	0.83 (0.62-1.09)	0.81 (0.61-1.07)	Ref.	1.11 (1.00-1.23)	1.12 (0.99-1.27)	Ref.	1.03 (0.79-1.35)	1.16 (0.89-1.51)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	0.92 (0.81-1.03)			1.05 (1.00-1.10)			1.00 (0.89-1.12)		
	P-trend	0.151			0.066			0.971		
	HR (95% CI)	Ref.	0.83 (0.63-1.10)	0.81 (0.60-1.07)	Ref.	1.11 (1.00-1.23)	1.12 (0.99-1.27)	Ref.	1.04 (0.80-1.35)	1.18 (0.90-1.54)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	0.91 (0.81-1.03)			1.05 (1.00-1.10)			1.01 (0.90-1.14)		
	P-trend	0.149			0.067			0.842		

Supplementary Table 5-2_continued *Risk of Bladder Cancer According to Intake of Individual Types of Vegetable*

Subgroup		Broad Bean (g/day)			French Bean (g/day)			Sweet Corn (g/day)		
Intake Tertiles		Low (0)	Medium (0-5)	High (>5)	Low (0-5)	Medium (5-13)	High (>13)	Low (0-2)	Medium (2-7)	High (>7)
No. Case		327	232	318	216	370	637	240	219	178
	HR (95% CI)	Ref.	1.23 (1.04-1.46)	1.24 (1.06-1.44)	Ref.	0.91 (0.77-1.09)	0.94 (0.79-1.11)	Ref.	1.17 (0.97-1.41)	1.10 (0.90-1.35)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	1.06 (1.00-1.13)			1.00 (0.94-1.06)			1.01 (0.93-1.10)		
	P-trend	0.036			0.957			0.851		
	HR (95% CI)	Ref.	1.23 (1.04-1.46)	1.23 (1.05-1.44)	Ref.	0.91 (0.77-1.09)	0.93 (0.79-1.11)	Ref.	1.18 (0.98-1.43)	1.11 (0.90-1.36)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	1.06 (1.00-1.13)			1.00 (0.94-1.06)			1.01 (0.93-1.10)		
	P-trend	0.037			0.984			0.825		

Supplementary Table 5-2_continued *Risk of Bladder Cancer According to Intake of Individual Types of Vegetable*

Subgroup		Okra (g/day)			Avocado (g/day)		
Intake Tertiles		Low (0)	Medium (0-4)	High (>4)	Low (0)	Medium (0-2)	High (>2)
No. Case		143	105	98	319	141	177
	HR (95% CI)	Ref.	1.15 (0.89-1.48)	0.96 0.74-1.25)	Ref.	1.01 (0.82-1.23)	1.04 (0.86-1.26)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	1.02 (0.91-1.15)			1.00 (0.92-1.08)		
	P-trend	0.698			0.977		
	HR (95% CI)	Ref.	1.17 (0.91-1.51)	1.01 (0.77-1.31)	Ref.	1.01 (0.83-1.24)	1.07 (0.88-1.29)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	1.05 (0.93-1.17)			1.01 0.93-1.09)		
	P-trend	0.447			0.839		

Supplementary Table 5-2_continued *Risk of Bladder Cancer According to Intake of Individual Types of Vegetable*

Subgroup		Olive (g/day)			Edible Fungi (g/day)		
Intake Tertiles		Low (0-0.4)	Medium (0.4-2)	High (>2)	Low (0-1)	Medium (1-5)	High (>5)
No. Case		96	89	106	1,040	1,058	768
	HR (95% CI)	Ref.	1.10 (0.82-1.48)	1.21 (0.91-1.61)	Ref.	1.02 (0.93-1.11)	1.02 (0.91-1.14)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	1.05 (0.95-1.15)			0.99 (0.94-1.04)		
	P-trend	0.335			0.626		
	HR (95% CI)	Ref.	1.11 (0.83-1.49)	1.22 (0.91-1.63)	Ref.	1.01 (0.93-1.12)	1.02 (0.91-1.14)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	1.05 (0.95-1.15)			0.99 (0.94-1.04)		
	P-trend	0.364			0.645		

[d] Model 1: Adjusted for age (years, continuous), sex (male or female, if applicable), smoking (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), and total energy intake (kcal/day, continuous).

[d, e] Model 2: Additionally, ethnicity (Caucasian or non-Caucasian, if applicable) , alcohol intake (ml/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), and grain intake (g/day, continuous).

Abbreviation: HR, hazard ratio; SD, standard deviation; CI, confidence interval; g, gram; kcal, kilocalorie.

The intervals of tertiles were categorized as open lower values and ended upper values.

Reference group was lowest intake (tertile 1).

P-trend <0.05 was considered statistically significant.

Supplementary Table 5-3 *Risk of Bladder Cancer According to Intake of Total and Subgroup Vegetable (After Removing 358 Bladder Cancer Cases Diagnosed within 2 years)*

Subgroup		Total Vegetable (g/day)			Total Vegetable for Females (g/day)			Non-starchy Vegetable (g/day)		
Intake Tertiles		Low (0-177)	Medium (117-221)	High (>221)	Low (0-177)	Medium (117-221)	High (>221)	Low (0-108)	Medium (108-211)	High (>211)
No. Case		861	864	1,170	255	283	270	757	720	1,377
	HR (95% CI)	Ref.	1.00 (0.90-1.10)	0.96 (0.85-1.08)	Ref.	0.88 (0.73-1.06)	0.78 (0.62-0.96)	Ref.	0.99 (0.89-1.11)	0.97 (0.85-1.11)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	0.99 (0.93-1.03)			0.88 (0.79-0.98)			0.98 (0.93-1.03)		
	P-trend	0.326			0.023			0.701		
	HR (95% CI)	Ref.	1.00 (0.90-1.10)	0.81 (0.70-0.93)	Ref.	0.88 (0.73-1.07)	0.78 (0.62-0.97)	Ref.	0.99 (0.89-1.11)	0.97 (0.85-1.11)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	0.97 (0.92-1.02)			0.88 (0.79-0.98)			0.97 (0.92-1.02)		
	P-trend	0.233			0.026			0.274		

Supplementary Table 5-3_continued *Risk of Bladder Cancer According to Intake of Total and Subgroup Vegetable (After Removing 358 Bladder Cancer Cases Diagnosed within 2 years)*

Subgroup		Non-starchy Vegetable for Females (g/day)			Starchy Vegetable (g/day)		
Intake Tertiles		Low (0-108)	Medium (108-211)	High (>211)	Low (0-7)	Medium (7-18)	High (>18)
No. Case		256	293	259	619	429	294
	HR (95% CI)	Ref.	0.88 (0.73-1.07)	0.77 (0.61-0.96)	Ref.	0.92 (0.81-1.04)	0.92 (0.79-1.07)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	0.88 (0.79-0.98)			0.95 (0.89-1.02)		
	P-trend	0.027			0.202		
	HR (95% CI)	Ref.	0.89 (0.73-1.07)	0.77 (0.61-0.96)	Ref.	0.93 (0.82-1.05)	0.93 (0.80-1.09)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	0.88 (0.79-0.99)			0.96 (0.89-1.03)		
	P-trend	0.026			0.248		

Supplementary Table 5-3_continued *Risk of Bladder Cancer According to Intake of Total and Subgroup Vegetable (After Removing 358 Bladder Cancer Cases Diagnosed within 2 years)*

Subgroup		Green Leafy Vegetable (g/day)			Cruciferous Vegetable (g/day)		
Intake Tertiles		Low (0-5)	Medium (5-23)	High (>23)	Low (0-8)	Medium (8-26)	High (>26)
No. Case		805	945	1,104	777	769	1,308
	HR (95% CI)	Ref.	1.01 (0.91-1.11)	0.93 (0.82-1.06)	Ref.	1.01 (0.91-1.11)	0.93 (0.82-1.06)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	0.99 (0.93-1.04)			0.97 (0.93-1.02)		
	P-trend	0.648			0.278		
	HR (95% CI)	Ref.	1.01 (0.92-1.12)	0.93 (0.82-1.06)	Ref.	0.96 (0.87-1.08)	0.93 (0.82-1.06)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	0.99 (0.93-1.04)			0.97 (0.93-1.02)		
	P-trend	0.631			0.262		

[d] Model 1: Adjusted for age (years, continuous), sex (male or female, if applicable), smoking (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), and total energy intake (kcal/day, continuous).

[d, e] Model 2: Additionally, ethnicity (Caucasian or non-Caucasian, if applicable) , alcohol intake (ml/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), and grain intake (g/day, continuous).

Abbreviation: HR, hazard ratio; SD, standard deviation; CI, confidence interval; g, gram; kcal, kilocalorie.

The intervals of tertiles were categorized as open lower values and ended upper values.

Reference group was lowest intake (tertile 1).

P-trend <0.05 was considered statistically significant.

Supplementary Table 5-4 Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Intake of Vegetable Stratified by Study Centre

Study Subgroup	Model Adjustments	Vegetable Intake (g/day)			P_{trend}
		Low	Medium	High	
NLCS	No. Cases	261	195	321	
	Model 2 [d, e]	Ref.	0.98 (0.83-1.16)	0.93 (0.78-1.11)	0.329
RERF-LSS	No. Cases	26	13	2	
	Model 2 [d, e]	Ref.	0.65 (0.33-1.29)	0.97 (0.23-4.13)	0.423
VITAL	No. Cases	121	124	101	
	Model 2 [d, e]	Ref.	1.09 (0.85-1.41)	1.09 (0.82-1.46)	0.389
MCCS	No. Cases	101	103	87	
	Model 2 [d, e]	Ref.	1.27 (0.96-1.69)	1.11 (0.82-1.50)	0.170
EPIC	No. Cases	717	575	406	
	Model 2 [d, e]	Ref.	0.91 (0.82-1.02)	0.69 (0.61-0.79)	<0.001

[d,e] Model 2: Adjusted for age (years, continuous), sex (male or female, if applicable), smoking (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), and total energy intake (kcal/day, continuous), ethnicity (Caucasian or non-

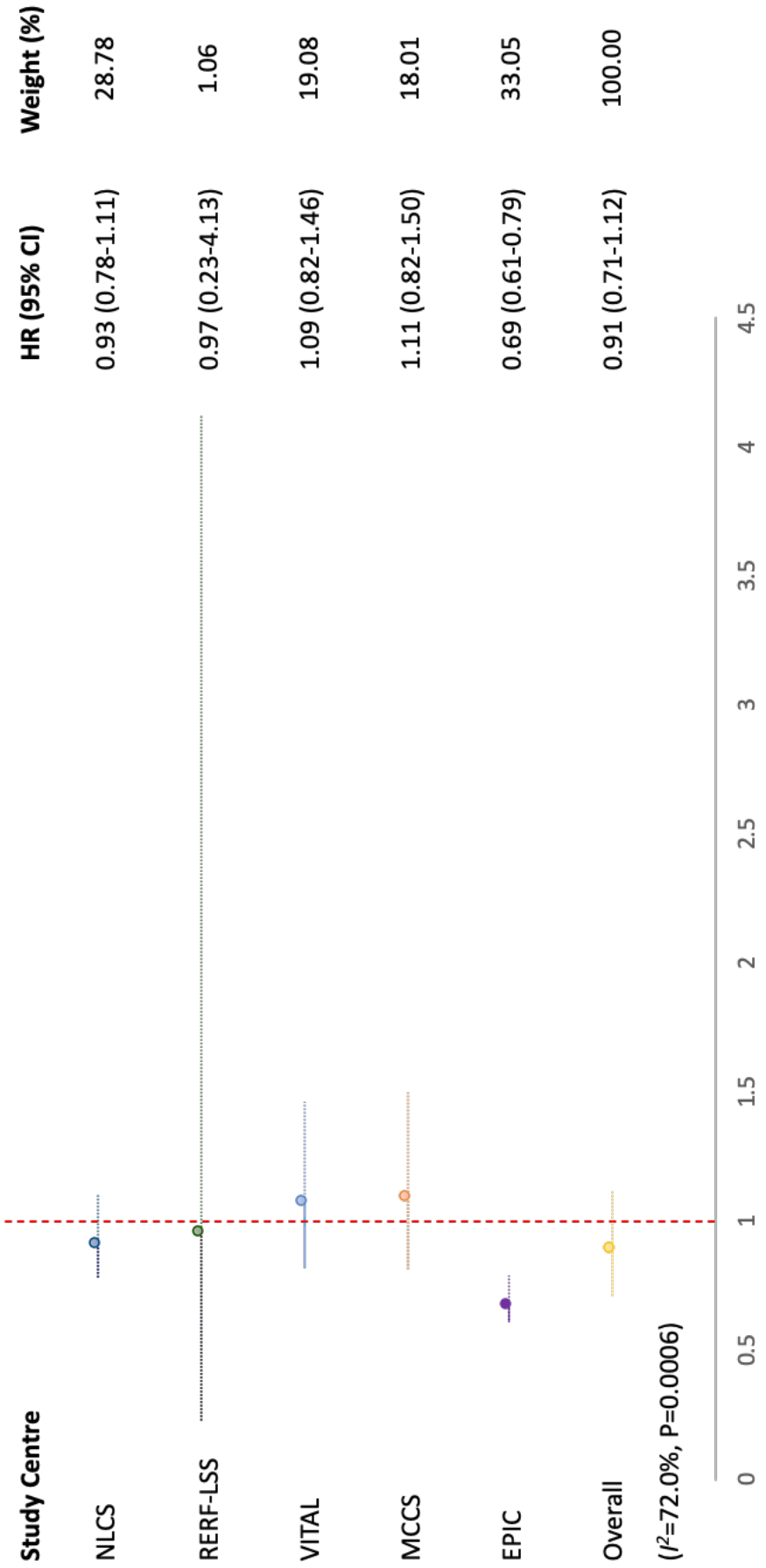
Caucasian, if applicable) , alcohol intake (ml/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), and grain intake (g/day, continuous).

Abbreviation: HR, hazard ratio; SD, standard deviation; CI, confidence interval; g, gram; kcal, kilocalorie.

The intervals of tertiles were categorized as open lower values and ended upper values.

Reference group was lowest intake (tertile 1).

P-trend <0.05 was considered statistically significant.


Study Centre
HR (95% CI)
Weight (%)
NLCS
0.93 (0.78-1.11)
28.78
RERF-LSS
0.97 (0.23-4.13)
1.06
VITAL
1.09 (0.82-1.46)
19.08
MCCS
1.11 (0.82-1.50)
18.01
EPIC
0.69 (0.61-0.79)
33.05
Overall
0.91 (0.71-1.12)
100.00
(I^2=72.0%, P=0.0006)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5

Supplementary Figure 5-1 *Forest Plot of Meta-Analyses with HRs and 95% CIs for High vs. Low Intake of Total Vegetable with Bladder Cancer Risk*

Diamond dots denote the hazard ratios (HRs); Horizontal lines represent the 95% confidence intervals (CIs); Weights (grey squares) are from random effects analyses.

Adjusted for age (years, continuous), sex (male or female, if applicable), smoking (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), and total energy intake (kcal/day, continuous), ethnicity (Caucasian or non-Caucasian, if applicable) , alcohol intake (ml/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), and grain intake (g/day, continuous).

Abbreviation: HR, hazard ratio; CI, confidence interval; g, gram; kcal, kilocalorie; ml, millilitre.

Reference group was lowest intake (tertile 1).

P-value <0.05 was considered statistically heterogeneous.

Chapter 6

Grain and Dietary Fibre Intake and Bladder Cancer Risk: A Pooled Analysis of 574,726 Participants from 13 Prospective Cohort Studies

Evan Y.W. Yu, Anke Wesselius, Siamak Mehrkanoon, Maree Brinkman, Piet van den Brandt, Emily White, Elisabete Weiderpass, Florence Le Calvez-Kelm, Marc Gunter, Inge Huybrechts, Fredrik Liedberg, Marko Lukic, Anne Tjonneland, Elio Riboli, Graham G. Giles, Roger L. Milne, Maurice P. Zeegers

Abstract

Objective: Higher intakes of whole grains and dietary fibre have been associated with lower risk of insulin resistance, hyperinsulinemia, and inflammation, which are known predisposing factors for cancer. As the evidence of association with bladder cancer is limited. We aimed to assess associations with bladder cancer risk for intakes of whole grains, refined grains and dietary fibre.

Methods: We pooled individual data from 574,726 participants in 13 cohort studies, 3,214 of whom developed incident bladder cancer. Pooled *hazard ratios* (HRs), with corresponding 95% *confidence intervals* (CIs), were estimated using Cox regression models stratified on cohort. Dose-response relationships were examined using fractional polynomial regression models.

Results: We found that higher intake of total whole grain was associated with lower risk of bladder cancer (comparing highest with lowest intake tertile: HR=0.87, 95% CI=0.77-0.98, P=0.021 for trend; HR per 1 SD increment=0.95, 95% CI=0.91-0.99). No association was observed for intake of total refined grain. Intake of total dietary fibre intake was also inversely associated with bladder cancer risk (comparing highest with lowest intake tertile: HR=0.86; 95% CI=0.74-0.98; P=0.021 for trend; HR per 1 SD increment=0.90, 95% CI=0.82-0.98). In addition, dose-response analyses gave estimated HRs of 0.97 (95% CI 0.95-0.99) for intake of total whole grain and 0.96 (95% CI 0.94-0.98) for intake of total dietary fibre per 5 grams daily increment. When considered jointly, highest intake of whole grains with the highest intake of dietary fibre showed 27% reduced risk (95% CI 0.56-0.95, P=0.034 for trend) of bladder cancer than lowest intake, suggesting potential synergism.

Conclusion: Higher intakes of total whole grain and total dietary fibre are associated with reduced risk of bladder cancer individually and jointly. Further studies are needed to clarify the underlying mechanisms for these findings.

Key Words: Bladder Cancer; Grain; Dietary Fibre; Dose-response Analysis; Cohort Study

Introduction

Bladder cancer is the 10th most common malignancy worldwide, with an estimated 550,000 new cases and 200,000 deaths annually [1, 2]. Incidence rates of bladder cancer are highest in Europe and North America, with a strong predominance in males and the elderly [3-8]. Bladder cancer is reported to be the most expensive of all cancer in terms of life time treatment due to its high rate of recurrence [9]. Diet has been suspected to be important, in addition to smoking and occupational exposure, but only arsenic-contaminated food is considered to be an established dietary risk factor for bladder cancer [10-14]. Because grain intake is an important component of numerous dietary guidelines globally, interest in the health effects of grain intake is increasing [15, 16].

Whole grains contain all components of the kernel, *i.e.*, the bran, germ, and endosperm. Both the bran outer coating and the inner germ are major sources of dietary fibre, vitamins, minerals, phytonutrients, and numerous other nutrients which may be beneficial to health [17]. However, during refining process, the outer bran and inner germ are removed and only the endosperm retained. This results in a substantial reduction in dietary fibre, vitamins, minerals, and other components. Although many vitamins and minerals are often added back to refined grains by subsequent process, the fibre content remains greatly diminished [18, 19].

An accumulation of evidence shows that intake of dietary fibre is associated with lower risk of insulin resistance, hyperinsulinemia [20], and inflammation [21], which are known predisposing factors for cancer; however, evidence of association with bladder cancer risk is sparse, with only two case-control studies reporting insufficient evidence of inverse association for intake of whole grains [22, 23]. In contrast to the beneficial health association of whole grains containing rich fibre, studies of refined grains mainly show no association with health [24-28], or harmful associations [29, 30] and there is no strong evidence of association with bladder cancer risk.

We therefore assessed association with the bladder cancer risk for intake of whole grains and refined grains, using data from 13 prospective cohort studies pooled in the *BLadder cancer Epidemiology and Nutritional Determinants* (BLEND) international study. In addition, we also

investigated the potential association of dietary fibre intake with bladder cancer risk by evaluating total and individual food sources (*i.e.* cereal, fruit, and vegetable fibre).

Methods

Study Sample

Data were obtained from BLEND, an international nutritional consortium currently consisting of 19 case-control studies and 16 cohort studies. Thirteen cohort studies with a total of 574,726 participants, 3,214 of whom developed incident bladder cancer, had sufficient information on grain intake to be eligible for inclusion in the present study. These studies originated from 12 countries in 3 continents [*i.e.* Europe: *European Prospective Investigation into Cancer and Nutrition* cohort study (EPIC) [31] (Denmark [32], France [33], Germany [34], Greece [35], Italy [36], Spain [35], Sweden [37, 38], The Netherlands [39], United Kingdom [40, 41], Norway [42]), *NetherLands Cohort Study* (NLCS) [43]; North America: *VITamins And Lifestyle* cohort study (VITAL) [44]; and Oceania: *Melbourne Collaborative Cohort Study* (MCCS) in Australia [45, 46]]. Person-years of follow-up for each participant were calculated from the date of study enrolment until the date of bladder cancer diagnosis, or date of last follow-up (*e.g.* date of death, lost to follow-up, or study exit), whichever came first. For the NLCS study, a nested case-cohort design was applied in order to increase the follow-up coverage and efficiency, in which the number of person-years at risk was estimated based on a sub-cohort that was randomly sampled [43]. Each study was approved by their local ethical research committee [31, 43, 44, 46] (**Supplementary Table 6-1**).

Data Collection and Coding

Details on the methodology of the BLEND consortium have been described elsewhere [47]. In brief, all included studies used self-administered or trained interviewer administered *food frequency questionnaire* (FFQ) that was validated on either food groups [44, 48-51], and/or energy intake [48, 51, 52]. For each study, participants were asked to report on their usual intake during the year before study enrolment of individual types of whole grains [*i.e.* brown rice, wheat or oat, and basic products of other cereals (*e.g.* buckwheat, millet, sorghum, or spelt)] and of refined grains [*i.e.* white rice, pasta or noodles, leavened bread, unleavened bread, other bakery wares, savoury cereal dishes (*e.g.* dumplings, couscous, risotto, pizza, pancake or pie), and breakfast cereals]. These data were harmonized using the hierarchal

Eurocode 2 food coding system developed by the European Union [53], with weekly, monthly or yearly intake converted to *grams* (g) per day. In order to extract dry weight (*e.g.* uncooked pasta or noodles, uncooked rice, uncooked wheat or oat) across all grains, the water content of grain was determined according to the composition database from the *United States Department of Agriculture* (USDA) and subtracted from the grains intake [54]. Total intake of dietary fibre and dietary fibre from cereal, fruit, and vegetable, were calculated by multiplying the amount of each food consumed by the dietary fibre content per gram according to USDA.

Each study ascertained incident bladder cancer, defined to include all urinary bladder neoplasms according to the *International Classification of Diseases for Oncology* (ICD-O-3 code C67) using population-based cancer registries, health insurance records, or medical records. Bladder cancers were classified as non-muscle invasive bladder cancer (NMIBC) or muscle invasive bladder cancer (MIBC). NMIBC included non-invasive papillary carcinomas confined to the urothelium (stage Ta), and carcinomas that invaded the lamina propria of the bladder wall (stage T1). High grade flat non-invasive carcinomas confined to the urothelium (*carcinoma in situ*; CIS) without other concomitant tumour stages [*i.e.* T1/Ta (classified to non-muscle invasive prior), or muscle invasive] were also classified as NMIBC. MIBC included carcinomas that invaded into the detrusor muscle (stage T2), carcinomas that invaded into the peri-vesical tissue (stage T3), and carcinomas that invaded adjacent tissues and organs (most often the prostate or uterus, stage T4).

In addition to information on grain and other dietary intakes, the BLEND dataset also included data on study characteristics (design, method of dietary assessment, geographical region), participant demographics (age, sex and ethnicity), smoking status and smoking pack-years (*i.e.* the number of cigarettes smoked per day multiplying the years of smoking), which were measured at baseline.

Statistical Analyses

To assess the influence of intake of grains and fibre on bladder cancer risk, Cox regression analyses with a stratification approach to adjust for cross-cohort heterogeneity [55], were used to estimate the pooled *hazard ratios* (HRs) and 95% *confidence intervals* (CIs). The proportional hazard assumption was examined for each analysis and no evidence of violation

was found. Additionally, the appropriateness of the use of the log-normal distribution was tested using a Wald test, and again, no evidence of violation was found. Grains (*i.e*. total grain, total whole grain, total refined grain, brown rice, wheat or oat, basic products of other cereals, white rice, pasta or noodles, leavened bread, unleavened bread, bakery wares, savoury cereals, and breakfast cereals) and dietary fibre intake (*i.e.* total dietary fibre from all food sources, cereal fibre, fruit fibre and vegetable fibre) were divided into 3 groups defined by tertile: low intake (tertile 1), medium intake (tertile 2) and high intake (tertile 3). Low intake was used as the reference group and associations were assessed applying two models, model 1) adjusted for age (years), sex (male and female), smoking, total energy intake [*kilocalorie* (kcal)/day, continuous; using residual model to remove extraneous variation [56]] and included cohort as a stratification variable and model 2) additionally, adjusted for ethnicity (Caucasian or non-Caucasian) and for potential dietary factors that affect the development of bladder cancer [10], including; alcohol intake [*millilitre* (ml)/day, continuous], sugar intake [*gram* (g)/day, continuous], meat intake (g/day, continuous), vegetable intake (g/day, continuous), fruit intake (g/day, continuous), and fat intake (g/day, continuous. Smoking was defined as a dummy variable: 0 (never smokers); 1 [current light smokers (*i.e.* smoking less than 20 pack-years)]; 2 [current heavy smokers (*i.e.* smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (*i.e.* smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (*i.e.* smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]. Stratified analyses were performed by bladder cancer subtype (*i.e.* NMIBC and MIBC) and sex. In addition, the HRs and 95% CIs of bladder cancer with per 1 *standard deviation* (SD) increase in grains and dietary fibre intake were also estimated using the same models. Furthermore, a potential joint association of total whole grain and total dietary fibre intake with bladder cancer risk was assessed using the lowest intake of both total whole grain and total dietary fibre as the reference. To test for linearity or nonlinearity, we included both linear and quadratic terms (*i.e.* the absolute intake and intake squared) in the models, then a likelihood ratio test was used to assess the difference between the nonlinear and linear models [57]. Since results showed no evidence of a non-linear association, linear models were applied in the present study. P for trend test was conducted by assigning medians to per 1 SD as a continuous variable in the models. The variables of bladder cancer

status (*i.e.* cases or non-cases), follow-up time, age, sex, smoking and total energy intake were complete without missing values. Missing variables (*e.g.* alcohol intake, sugar intake, meat intake, vegetable intake, fruit intake, and fat intake; missing proportion were all less than 5%) were imputed separately in each participating cohort by multiple imputation method. Only participants with complete information on bladder cancer status, age, sex, smoking and total energy intake were included when building the imputation models. Linear regression models were then fitted for those variables with missing data separately.

In our secondary analyses, potential dose-response relationships of total whole grain, total refined grain, total dietary fibre and cereal fibre with bladder cancer risk were assessed by using fractional polynomial regression from the natural logarithm of the HRs across categories of intake, in which the best fitting second order fractional polynomial regression model was defined as the model with the lowest deviance [58, 59]. For this, we categorized each source of grain (*e.g.* total whole grain or total refined grain) and dietary fibre (*e.g.* total dietary fibre or cereal fibre) into 10 doses according to the range of each grain or dietary fibre intake, by which the intervals of each intake were different. P value for trend were estimated by assigning medians to each category of intake as a continuous variable. A likelihood ratio test was used to assess the difference between the nonlinear (*i.e.* the absolute dose and dose squared) and linear (*i.e.* the absolute dose) models test for linearity or nonlinearity [57]. Model 2 was applied in dose-responses analyses.

A sensitivity analysis was performed by removing cases diagnosed within the first 2 years after recruitment to each study. An extra analysis for total refined grain was assessed by excluding pasta source in order to test whether the possible misclassification of pasta would influence the result. In addition, the role of smoking was furtherly tested by replacing the smoking dummy variable by both smoking status (never, former and current) and smoking pack-years (continuous). As a last step, the association between intake of total whole grain, total refined grain and total dietary fibre was assessed is each participating cohort separately and combined in a meta-analysis approach using a random-effect model.

All statistical analyses were performed using STATA version 14 SE (Stata Corporation, Texas, USA). A two-tailed P values <0.05 was considered statistically significant.

Results

Baseline Characteristics

The baseline characteristics of the study sample are shown in **Table 6-1**. In total, 574,726 study participants contributed 6,335,667 person-years of follow-up over a median of 11 years, with 3,214 incident bladder cancer cases (2,416 males, 798 females) diagnosed. Of these, 2,041 (63%) cases had available diagnosis records of NMIBC (39%) or MIBC (24%). The median age at baseline was 53 years. The majority (99.3%) of participants were Caucasian. Those with higher intake of total whole grain and total dietary fibre were less likely to be smokers and more likely to be males. Those with higher intake of total refined grain were more frequently current smokers, males and of younger age.

Associations of Grains and Dietary Fibre intake with Bladder Cancer Risk

Total grain intake and bladder cancer risk

For the intake of "total grain", no evidence of association was observed overall, by cancer subtype, or by sex (**Table 6-2**).

Whole grains intake and bladder cancer risk

The results of the Cox regression analyses for the associations between total whole grain and bladder cancer risk are shown in **Table 6-2**. In multivariable-adjusted analyses (model 2), higher total whole grain intake was significantly associated with lower bladder cancer risk (comparing the highest to the lowest tertile intake: HR=0.87, 95% CI=0.77-0.98, P=0.021 for trend; HR per 1 SD increment=0.95, 95% CI=0.91-0.99). However, no evidence of association was observed in the stratified analyses by cancer subtype. Results were consistent for both males (comparing the highest with lowest intake tertile: HR_{model2}=0.83, 95% CI=0.71-0.97, P=0.051 for trend; HR per 1 SD increment=0.95, 95% CI=0.90-1.00) and females (comparing the highest with lowest intake tertile: HR_{model2}=0.83, 95% CI=0.71-0.96, P=0.053 for trend; HR per 1 SD increment=0.93, 95% CI=0.85-1.01).

Of the individual whole grain assessed, only higher intake of brown rice was significantly associated with a decreased bladder cancer risk (comparing the highest with lowest intake tertile: HR_{model2}= 0.78, 95% CI=0.67-0.92, P=0.001 for trend, HR per 1 SD increment=0.89, 95% CI=0.82-0.95) (**Table 6-3**). All other whole grains showed a null association.

Refined grains intake and bladder cancer risk

No evidence of association between total refined grain intake and bladder cancer risk was observed (**Table 6-2**). Looking at the individual refined grain sources, similar null-associations were found, except for the intake of "pasta or noodles", which showed to be borderline inversely associated with bladder cancer risk when comparing medium intake versus low intake (HR_{model2}=0.90, 95% CI=0.81-0.99, P=0.649 for trend; HR per 1 SD increment=0.99, 95% CI=0.94-1.04) (**Table 6-3**).

Dietary fibre intake and bladder cancer risk

The associations between the intake of total dietary fibre and dietary fibre from different food sources and bladder cancer risk are shown in **Table 6-4**. The intake of total dietary fibre was inversely associated with bladder cancer risk (comparing the highest with lowest intake tertile: HR_{model2}=0.86; 95% CI=0.74-0.98; P=0.021 for trend; HR per 1 SD increment=0.90, 95% CI=0.82-0.98). No evidence of association was observed for any of the individual dietary fibre food sources.

Joint association of total whole grain and total dietary fibre with bladder cancer risk

The results of the Cox regression analyses for the potential joint effect of total whole grain and total dietary fibre on bladder cancer are shown in **Table 6-5**. Individuals with the highest intake of both total whole grain and total dietary fibre showed a 27% reduced bladder cancer risk (95% CI 0.56-0.95, P=0.034 for trend) compared to individuals with the lowest intake of both total whole grain and total dietary fibre.

Table 6-1 *Characteristics of the Study Population (3,214 Cases and 571,512 Non-cases) of Grains and Fibre Intake of Included Studies of Included Studies*

		Overall Population (%)	TMN Stage			Sex	
			Non-cases (%)	Cases (%)	MIBC (%)	Male (%)	Female (%)
Total Whole Grain (g/day)	Low (0-3)	72,821 (33.5)	71,830 (33.3)	991 (57.2)	360 (63.7)	22,476 (36.1)	50,345 (32.4)
	Medium (3-8)	74,285 (34.1)	73,932 (34.3)	353 (20.4)	92 (16.3)	18,167 (29.2)	56,118 (36.2)
	High (>8)	70,450 (32.4)	70,061 (32.4)	389 (22.4)	113 (20.0)	21,659 (34.7)	48,791 (31.4)
	Mean (SD)	9.82 (12.69)	10.79 (14.19)	9.81 (12.68)	5.73 (11.50)	10.98 (14.52)	9.40 (11.93)
Total Refined Grain (g/day)	Low (0-102)	191,576 (33.4)	190,572 (33.4)	1,004 (31.3)	267 (34.5)	57,331 (30.4)	134,245 (34.8)
	Medium (102-181)	191,575 (33.3)	190,337 (33.3)	1,238 (38.5)	356 (46.1)	51,474 (27.3)	140,101 (36.3)
	High (>181)	191,575 (33.3)	190,603 (33.4)	972 (30.2)	150 (19.4)	80,057 (42.4)	111,518 (28.9)
	Mean (SD)	156.37 (102.19)	156.39 (102.24)	153.34 (93.32)	135.49 (75.98)	177.41 (118.64)	146.08 (91.34)
Total Dietary Fibre (g/day)	Low (0-17)	191,576 (33.4)	190,402 (33.3)	1,174 (36.5)	273 (35.3)	67,336 (35.7)	124,240 (32.2)
	Medium (17-25)	191,575 (33.3)	190,371 (33.3)	1,204 (37.5)	323 (41.8)	59,435 (31.5)	132,140 (34.3)
	High (>25)	191,575 (33.3)	190,739 (33.4)	836 (26.0)	177 (22.9)	62,091 (32.8)	129,484 (33.6)
	Mean (SD)	23.37 (12.38)	23.38 (12.40)	21.04 (9.07)	20.68 (8.56)	22.02 (10.68)	24.03 (13.09)

Table 6-1_continued *Characteristics of the Study Population (3,214 Cases and 571,512 Non-cases) of Grains and Fibre Intake of Included Studies*

		Pearson Years	Total Energy (kcal/day)	Age [a]				
				years	≤55	55-60	60-65	65-70
Total Whole Grain (g/day)	Low (0-3)	1,016,724	2,055.90	52.98	40,336 (28.4)	12,111 (37.8)	10,241 (41.8)	6,919 (52.1)
	Medium (3-8)	870,317	2,074.74	49.75	50,475 (35.6)	10,592 (33.1)	7,687 (31.4)	3,653 (27.5)
	High (>8)	855,859	2,251.09	48.51	50,994 (36.0)	9,300 (29.1)	6,557 (26.8)	2,709 (20.4)
	Mean (SD)	9.82 (12.69)	9.82 (12.69)	9.82 (12.69)	10.22 (12.66)	9.04 (11.61)	9.13 (12.47)	8.92 (13.95)
Total Refined Grain (g/day)	Low (0-102)	1,929,466	1,731.56	54.96	97,204 (28.6)	33,855 (34.5)	28,094 (35.4)	18,125 (50.6)
	Medium (102-181)	2,262,417	2,002.27	52.07	117,384 (34.5)	31,659 (32.3)	26,446 (33.3)	11,241 (31.4)
	High (>181)	2,143,784	2,455.98	50.90	125,269	32,527 (33.2)	24,846 (31.3)	6,464 (18.0)
	Mean (SD)	156.38 (102.23)	156.37 (102.19)	156.37 (102.19	166.82 (104.19)	154.42 (102.14)	149.48 (96.97)	116.61 (83.20)
Total Dietary Fibre (g/day)	Low (0-17)	2,001,790	1,726.74	54.93	97,474 (28.7)	33,998 (34.7)	29,037 (36.6)	17,708 (49.4)
	Medium (17-25)	2,187,860	2,016.21	51.86	117,242 (34.5)	32,299 (32.9)	26,494 (33.4)	10,284 (28.7)
	High (>25)	2,146,058	2,446.86	51.14	125,141	31,744 (32.4)	23,855 (30.0)	7,838 (21.9)
	Mean (SD)	23.37 (12.38)	23.37 (12.38)	23.37 (12.38)	24.52 (12.49)	23.03 (12.43)	22.34 (11.85)	19.74 (11.78)

Table 6-1_continued *Characteristics of the Study Population (3,214 Cases and 571,512 Non-cases) of Grains and Fibre Intake of Included Studies*

		Age [a] (years)		Smoking Status			Smoking Pack-years	Smoking [c]
		70-75	>75	Never (%)	Former (%)	Current (%)		0 (%)
Total Whole Grain (g/day)	Low (0-3)	2,311 (54.7)	903 (51.4)	38,321 (32.0)	21,503 (36.3)	12,997 (33.7)	22.99	38,321 (32.0)
	Medium (3-8)	1,305 (30.9)	573 (32.6)	42,404 (35.4)	19,219 (32.4)	12,662 (32.9)	19.22	42,404 (35.4)
	High (>8)	610 (14.4)	280 (16.0)	39,024 (32.6)	18,553 (31.3)	12,873 (33.4)	19.72	39,024 (32.6)
	Mean (SD)	8.08 (12.60)	8.15 (13.10)	9.91 (12.91)	9.43 (12.97)	9.36 (11.52)	9.65 (12.30)	9.91 (12.91)
Total Refined Grain (g/day)	Low (0-102)	12,691 (68.5)	1,607 (51.9)	97,586 (34.1)	62,296 (36.8)	31,694 (26.6)	24.50	97,586 (34.1)
	Medium (102-181)	3,917 (22.2)	928 (30.0)	98,375 (34.4)	52,397 (31.0)	40,803 (34.3)	22.66	98,375 (34.4)
	High (>181)	1,909 (10.3)	560 (18.1)	90,409 (31.6)	54,517 (32.2)	46,649 (39.2)	23.08	90,409 (31.6)
	Mean (SD)	88.89 (68.98)	111.71 (76.90)	152.51 (98.36)	151.42 (103.60)	172.70 (107.43)	150.29 (102.03)	152.51 (98.36)
Total Dietary Fibre (g/day)	Low (0-17)	12,691 (68.5)	1,331 (43.0)	89,338 (31.2)	62,807 (37.1)	39,431 (33.1)	24.84	89,338 (31.2)
	Medium (17-25)	3,917 (22.2)	897 (29.0)	94,872 (33.1)	54,623 (32.3)	42,080 (35.3)	22.72	94,872 (33.1)
	High (>25)	1,909 (10.3)	867 (28.0)	102,160 (35.7)	51,780 (30.6)	37,635 (31.6)	22.36	102,160 (35.7)
	Mean (SD)	15.85 (8.61)	21.23 (11.68)	24.37 (13.21)	22.25 (11.83)	22.56 (10.79)	23.03 (12.67)	24.37 (13.21)

Table 6-1_continued *Characteristics of the Study Population (3,214 Cases and 571,512 Non-cases) of Grains and Fibre Intake of Included Studies*

		Smoking [c]						Ethnicity (%)
		1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)	Non-Caucasian
Total Whole Grain (g/day)	Low (0-3)	5,526 (29.8)	5,930 (39.7)	3,555 (66.6)	2,505 (72.0)	1,541 (30.4)	15,443 (30.6)	
	Medium (3-8)	6,513 (35.2)	4,418 (29.6)	693 (13.0)	368 (10.6)	1,731 (34.2)	18,158 (36.0)	
	High (>8)	6,489 (35.0)	4,594 (30.8)	1,087 (20.4)	606 (17.4)	1,790 (35.4)	16,860 (33.4)	
	Mean (SD)	9.55 (11.85)	8.95 (11.10)	11.99 (16.22)	12.53 (16.88)	9.82 (11.33)	9.73 (12.58)	
Total Refined Grain (g/day)	Low (0-102)	16,366 (28.9)	12,788 (24.8)	22,466 (81.6)	8,996 (74.8)	2,540 (23.3)	30,834 (23.8)	3,772 (94.0)
	Medium (102-181)	19,438 (34.3)	18,134 (35.1)	4,201 (15.3)	2,450 (20.4)	3,231 (29.7)	45,746 (35.3)	228 (5.7)
	High (>181)	20,834 (36.8)	20,703 (40.1)	869 (3.2)	582 (4.8)	5,112 (47.0)	53,066 (40.9)	15 (0.3)
	Mean (SD)	166.01 (104.24)	176.88 (109.31)	67.58 (46.92)	76.84 (52.01)	187.69 (112.19)	176.15 (103.30)	49.05 (31.28)
Total Dietary Fibre (g/day)	Low (0-17)	18,745 (33.1)	17,486 (33.9)	19,501 (70.8)	8,111 (67.4)	3,200 (29.4)	35,195 (27.1)	3,302 (82.2)
	Medium (17-25)	20,112 (35.5)	18,410 (35.7)	5,459 (19.8)	2,612 (21.7)	3,558 (32.7)	46,552 (35.9)	514 (12.8)
	High (>25)	17,781 (31.4)	15,729 (30.4)	2,576 (9.4)	1,305 (10.9)	4,125 (37.9)	47,899 (37.0)	199 (5.0)
	Mean (SD)	22.49 (10.71)	22.31 (10.46)	14.31 (8.10)	14.93 (8.31)	24.15 (12.48)	24.61 (11.77)	11.88 (7.26)

Table 6-1_continued *Characteristics of the Study Population (3,214 Cases and 571,512 Non-cases) of Grains and Fibre Intake of Included Studies*

		Ethnicity (%)	Alcohol (ml/day)	Fat (g/day)	Fruit (g/day)	Meat (g/day)	Sugar (g/day)	Vegetable (g/day)
		Caucasian						
Total Whole Grain (g/day)	Low (0-3)	72,821 (33.5)	9.17	24.21	85.93	85.93	22.10	222.97
	Medium (3-8)	74,285 (34.1)	9.79	22.77	82.24	82.24	23.97	239.05
	High (>8)	70,450 (32.4)	11.09	24.55	88.92	88.92	23.52	263.39
	Mean (SD)	9.82 (12.69)	9.82 (12.69)	9.82 (12.69)	9.82 (12.69)	9.82 (12.69)	9.82 (12.69)	9.82 (12.69)
Total Refined Grain (g/day)	Low (0-102)	187,804 (32.9)	7.20	16.88	54.48	54.48	20.73	181.77
	Medium (102-181)	191,347 (33.5)	9.57	25.30	75.45	75.45	23.32	212.32
	High (>181)	191,560 (33.6)	12.44	31.74	91.59	91.59	27.37	204.19
	Mean (SD)	157.13 (102.12)	156.37 (102.19)	156.37 (102.19)	156.37 (102.19)	156.37 (102.19)	156.37 (102.19)	156.37 (102.19)
Total Dietary Fibre (g/day)	Low (0-17)	188,274 (33.0)	8.17	18.55	54.50	54.50	18.69	130.47
	Medium (17-25)	191,061 (33.5)	9.77	25.78	73.94	73.94	20.77	189.46
	High (>25)	191,376 (33.5)	11.26	29.59	93.09	93.09	31.96	278.35
	Mean (SD)	23.45 (12.37)	23.37 (12.38)	23.37 (12.38)	23.37 (12.38)	23.37 (12.38)	23.37 (12.38)	23.37 (12.38)

[a] Age at the time of recruitment.

[b] Pack-years was defined as the number of cigarettes smoked per day multiplying the years of smoking.

[c] Smoking was defined as a dummy variable: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)].

Abbreviation: SD, standard deviation; TNM, tumour, nodes and metastasis classification; g, gram; ml, millilitre; kcal, kilocalorie; MIBC, muscle invasive bladder cancer; NMIBC, non-muscle invasive bladder cancer.

Table 6-2 *Risk of Bladder Cancer According to Intake of Total Grain, Total Whole Grain and Total Refined Grain*

Subgroup		Total Grains (g/day)								
		Overall			MIBC			NMIBC		
Intake Tertiles		Low (0-105)	Medium (105-186)	High (>186)	Low (0-105)	Medium (105-186)	High (>186)	Low (0-105)	Medium (105-186)	High (>186)
No. Case/ Participants		1,005/191,576	1,227/191,575	982/191,575	263/190,834	350/190,698	160/190,753	425/190,996	481/190,829	352/190,955
	HR (95% CI)	Ref.	0.93 (0.85-1.03)	0.95 (0.85-1.06)	Ref.	0.92 (0.76-1.11)	0.84 (0.66-1.08)	Ref.	0.97 (0.82-1.14)	0.95 (0.78-1.15)
Model 1 [d]	**HR Per 1 SD Increase (95% CI)**	0.97 (0.92-1.02)			0.93 (0.82-1.06)			0.99 (0.91-1.08)		
	P-trend	0.173			0.267			0.810		
	HR (95% CI)	Ref.	0.93 (0.84-1.02)	0.94 (0.83-1.05)	Ref.	0.90 (0.74-1.09)	0.80 (0.62-1.03)	Ref.	0.96 (0.82-1.13)	0.93 (0.76-1.13)
Model 2 [d, e]	**HR Per 1 SD Increase (95% CI)**	0.96 (0.91-1.01)			0.90 (0.79-1.04)			0.98 (0.90-1.07)		
	P-trend	0.124			0.145			0.705		

Table 6-2_continued *Risk of Bladder Cancer According to Intake of Total Grain, Total Whole Grain and Total Refined Grain*

Subgroup		Total Grains (g/day)						Total whole grain (g/day)		
		Male			Female			Overall		
Intake Tertiles		Low (0-113)	Medium (113-215)	High (>215)	Low (0-102)	Medium (102-173)	High (>173)	Low (0-3)	Medium (3-8)	High (>8)
No. Case/ Participants		793/62,954	1,049/62,954	574/62,954	287/128,622	291/128,621	220/128,621	991/72,821	353/74,285	389/70,450
	HR (95% CI)	Ref.	0.90 (0.80-1.01)	0.89 (0.78-1.01)	Ref.	0.98 (0.82-1.18)	1.12 (0.90-1.39)	Ref.	0.97 (0.85-1.11)	0.86 (0.76-0.97)
Model 1 [d]	**HR Per 1 SD Increase (95% CI)**	0.92 (0.86-0.99)			1.08 (0.98-1.20)			0.94 (0.90-0.98)		
	P-trend	0.015			0.118			0.013		
	HR (95% CI)	Ref.	0.89 (0.80-1.00)	0.88 (0.77-1.02)	Ref.	0.98 (0.81-1.18)	1.10 (0.88-1.38)	Ref.	1.01 (0.89-1.15)	0.87 (0.77-0.98)
Model 2 [d, e]	**HR Per 1 SD Increase (95% CI)**	0.91 (0.85-0.98)			1.08 (0.97-1.19)			0.95 (0.91-0.99)		
	P-trend	0.010			0.165			0.021		

Table 6-2_continued *Risk of Bladder Cancer According to Intake of Total Grain, Total Whole Grain and Total Refined Grain*

Subgroup		Total whole grain (g/day)								
		MIBC			NMIBC			Males		
Intake Tertiles		Low (0-3)	Medium (3-8)	High (>8)	Low (0-3)	Medium (3-8)	High (>8)	Low (0-3)	Medium (3-9)	High (>9)
No. Case/ Participants		360/72,190	92/74,024	113/70,174	424/72,254	133/72,065	156/70,217	787/22,476	259/19,149	295/20,677
	HR (95% CI)	Ref.	1.20 (0.95-1.51)	0.87 (0.70-1.08)	Ref.	1.07 (0.87-1.32)	0.84 (0.69-1.02)	Ref.	0.95 (0.82-1.10)	0.82 (0.71-0.95)
Model 1 [d]	**HR Per 1 SD Increase (95% CI)**	0.92 (0.85-0.99)			0.96 (0.90-1.02)			0.94 (0.89-0.99)		
	P-trend	0.033			0.210			0.027		
	HR (95% CI)	Ref.	1.20 (0.95-1.52)	0.87 (0.70-1.08)	Ref.	1.07 (0.87-1.32)	0.85 (0.70-1.03)	Ref.	0.98 (0.85-1.14)	0.83 (0.71-0.97)
Model 2 [d, e]	**HR Per 1 SD Increase (95% CI)**	0.92 (0.85-1.00)			0.96 (0.90-1.03)			0.95 (0.90-1.00)		
	P-trend	0.040			0.258			0.051		

Table 6-2_continued *Risk of Bladder Cancer According to Intake of Total Grain, Total Whole Grain and Total Refined Grain*

Subgroup		Total whole grain (g/day)			Total refined grains (g/day)					
		Females			Overall			MIBC		
Intake Tertiles		Low (0-10)	Medium (10-21)	High (>21)	Low (0-102)	Medium (102-181)	High (>181)	Low (0-102)	Medium (102-181)	High (>181)
No. Case/ Participants		204/51,754	104/51,830	84/51,670	1,004/191,576	1,238/191,575	972/191,575	267/190,839	356/190,693	150/190,753
	HR (95% CI)	Ref.	0.98 (0.84-1.13)	0.84 (0.73-0.97)	Ref.	0.94 (0.85-1.03)	0.96 (0.86-1.08)	Ref.	0.93 (0.77-1.12)	0.86 (0.67-1.11)
Model 1 [d]	**HR Per 1 SD Increase (95% CI)**	0.93 (0.86-1.00)			0.98 (0.93-1.03)			0.97 (0.85-1.10)		
	P-trend	0.048			0.334			0.612		
	HR (95% CI)	Ref.	0.98 (0.83-1.15)	0.83 (0.71-0.96)	Ref.	0.94 (0.85-1.03)	0.95 (0.84-1.07)	Ref.	0.90 (0.74-1.09)	0.81 (0.62-1.06)
Model 2 [d, e]	**HR Per 1 SD Increase (95% CI)**	0.93 (0.85-1.01)			0.97 (0.92-1.02)			0.94 (0.82-1.08)		
	P-trend	0.053			0.244			0.382		

Table 6-2_continued *Risk of Bladder Cancer According to Intake of Total Grain, Total Whole Grain and Total Refined Grain*

Subgroup		Total whole grain (g/day)								
		NMIBC			Male			Female		
Intake Tertiles		Low (0-102)	Medium (102-181)	High (>181)	Low (0-111)	Medium (111-211)	High (>211)	Low (0-99)	Medium (99-169)	High (>169)
No. Case/ Participants		422/190,994	491/190,828	355/190,958	808/62,954	1,048/62,954	560/62,954	283/128,623	295/128,620	220/128,621
	HR (95% CI)	Ref.	1.01 (0.86-1.18)	0.99 (0.82-1.21)	Ref.	0.90 (0.81-1.00)	0.89 (0.77-1.02)	Ref.	1.01 (0.84-1.21)	1.14 (0.91-1.48)
Model 1 [d]	**HR Per 1 SD Increase (95% CI)**	1.00 (0.92-1.09)			0.93 (0.87-1.00)			1.09 (0.99-1.21)		
	P-trend	0.997			0.038			0.087		
	HR (95% CI)	Ref.	1.00 (0.85-1.18)	0.97 (0.79-1.19)	Ref.	0.89 (0.79-1.00)	0.87 (0.74-1.01)	Ref.	1.01 (0.84-1.21)	1.12 (0.89-1.40)
Model 2 [d, e]	**HR Per 1 SD Increase (95% CI)**	0.99 (0.91-1.08)			0.93 (0.86-1.00)			1.08 (0.98-1.20)		
	P-trend	0.862			0.042			0.127		

[d] Model 1: Adjusted for age (years, continuous), sex (male or female, if applicable), smoking (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), and total energy intake (kcal/day, continuous).

[d, e] Model 2: Additionally, ethnicity (Caucasian or non-Caucasian, if applicable) , alcohol intake (ml/day, continuous), fruit intake (g/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), and vegetable intake (g/day, continuous).

Abbreviation: HR, hazard ratio; CI, confidence interval; SD, standard deviation; MIBC, muscle invasive bladder cancer; NMIBC, non-muscle invasive bladder cancer; g, gram; kcal, kilocalorie; ml, millilitre.

The intervals of tertiles were categorized as open lower values and ended upper values.

Reference group was lowest intake (tertile 1).

P-trend <0.05 was considered statistically significant.

Table 6-3 *Risk of Bladder Cancer According to Individual Intake of Whole Grains and Refined Grains*

Subgroup		Whole Grains (g/day)									Refined Grains (g/day)
		Brown Rice			Wheat or Oat			Basic Products of Other Cereals [f]			White Rice
Intake Tertiles		Low (0-4)	Medium (4-9)	High (>9)	Low (0)	Medium (0-2)	High (>2)	Low (0)	Medium (0-3)	High (>3)	Low (0-4)
No. Case/ Participants		910/64,959	262/64,685	270/64,822	877/15,715	81/3,590	210/9,032	820/4,802	25/233	32/212	976/44,980
Model 1 [d]	**HR (95% CI)**	Ref.	0.96 (0.82-1.11)	0.79 (0.68-0.92)	Ref.	1.21 (0.95-1.52)	0.92 (0.79-1.08)	Ref.	0.78 (0.52-1.16)	1.00 (0.70-1.42)	Ref.
	HR Per 1 SD Increase	0.92 (0.86-0.98)			0.98 (0.92-1.06)			0.98 (0.91-1.05)			0.95 (0.88-1.04)
	P-trend	0.010			0.644			0.555			0.269
Model 2 [d, e]	**HR (95% CI)**	Ref.	0.97 (0.83-1.13)	0.78 (0.67-0.92)	Ref.	1.20 (0.95-1.52)	0.93 (0.80-1.09)	Ref.	0.78 (0.52-1.17)	1.03 (0.72-1.47)	Ref.
	HR Per 1 SD Increase	0.89 (0.82-0.95)			0.99 (0.92-1.06)			0.98 (0.91-1.06)			0.96 (0.88-1.04)
	P-trend	0.001			0.725			0.635			0.344

Table 6-3_continued *Risk of Bladder Cancer According to Individual Intake of Whole Grains and Refined Grains*

Subgroup		Refined Grains (g/day)									
				Pasta or Noodles			Leavened Bread			Unleavened Bread	
Intake Tertiles		Medium (4-11)	High (>11)	Low (0-3)	Medium (3-9)	High (>9)	Low (0-73)	Medium (73-160)	High (>160)	Low (0)	Medium (0-4)
No. Case/ Participants		288/44,951	221/44,954	806/193,351	787/188,377	744/187,751	1,057/191,576	1,260/191,594	897/191,556	775/119,122	939/181,124
	HR (95% CI)	1.08 (0.92-1.26)	1.05 (0.91-1.21)	Ref.	0.89 (0.81-0.99)	0.89 (0.79-1.01)	Ref.	1.01 (0.91-1.11)	1.02 (0.90-1.15)	Ref.	0.94 (0.84-1.05)
Model 1 [d]	**HR Per 1 SD Increase**			0.99 (0.94-1.04)			1.00 (0.95-1.05)			0.95 (0.89-1.00)	
	P-trend			0.603			0.948			0.068	
	HR (95% CI)	1.09 (0.93-1.28)	1.05 (0.92-1.21)	Ref.	0.90 (0.81-0.99)	0.90 (0.80-1.01)	Ref.	0.99 (0.89-1.10)	1.02 (0.89-1.16)	Ref.	0.94 (0.84-1.06)
Model 2 [d, e]	**HR Per 1 SD Increase**			0.99 (0.94-1.04)			0.99 (0.94-1.05)			0.95 (0.89-1.00)	
	P-trend			0.649			0.768			0.069	

Table 6-3 continued *Risk of Bladder Cancer According to Individual Intake of Whole Grains and Refined Grains*

Subgroup		Refined Grains (g/day)									
		Bakery Wares				Savoury Cereals Dishes [g]			Breakfast Cereals		
Intake Tertiles		High (>4)	Low (0)	Medium (0-27)	High (>27)	Low (0-3)	Medium (3-7)	High (>7)	Low (0-6)	Medium (6-27)	High (>27)
No. Case/ Participants		863/181,124	1,732/477,213	688/14,011	448/14,011	161/28,872	96/18,996	89/21,623	1,013/33,151	251/32,949	250/31,728
	HR (95% CI)	0.97 (0.87-1.09)	Ref.	1.09 (0.76-1.55)	1.01 (0.70-1.45)	Ref.	0.94 (0.73-1.22)	0.86 (0.66-1.14)	Ref.	1.00 (0.85-1.17)	0.94 (0.79-1.13)
Model 1 [d]	**HR Per 1 SD Increase**		0.98 (0.96-1.01)			0.94 (0.82-1.07)			0.96 (0.89-1.03)		
	P-trend		0.267			0.323			0.280		
	HR (95% CI)	0.97 (0.87-1.09)	Ref.	1.09 (0.77-1.56)	1.01 (0.71-1.46)	Ref.	0.96 (0.74-1.24)	0.89 (0.67-1.17)	Ref.	1.01 (0.86-1.20)	0.97 (0.81-1.16)
Model 2 [d, e]	**HR Per 1 SD Increase**		0.98 (0.96-1.01)			0.95 (0.83-1.08)			0.97 (0.90-1.05)		
	P-trend		0.283			0.427			0.430		

[d] Model 1: Adjusted for age (years, continuous), sex (male or female), smoking (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), and total energy intake (kcal/day, continuous).

[d, e] Model 2: Additionally, ethnicity (Caucasian or non-Caucasian, if applicable) , alcohol intake (ml/day, continuous), fruit intake (g/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), and vegetable intake (g/day, continuous).

[f] "Basic products of other cereals": buckwheat, millet, sorghum, or spelt.

[g] "Savoury Cereals Dishes": dumplings, couscous, risotto, pizza, pancake or pie.

Abbreviation: HR, hazard ratio; CI, confidence interval; SD, standard deviation; g, gram; kcal, kilocalorie; ml, millilitre.

The intervals of tertiles were categorized as open lower values and ended upper values.

Reference group was lowest intake (tertile 1).

P-trend <0.05 was considered statistically significant.

Table 6-4 *Risk of Bladder Cancer According to Intake of Total Dietary Fibre and Individual Source of Dietary Fibre*

Subgroup		**Total Dietary fibre (g/day)**			**Cereal Fibre (g/day)**		
Intake Tertiles		Low (0-17)	Medium (17-25)	High (>25)	Low (0-7)	Medium (7-12)	High (>12)
No. Case/ Participants		1,015/191,576	1,097/191,575	1,102/191,575	1,111/191,576	1,203/191,576	900/191,574
	HR (95% CI)	Ref.	0.93 (0.84-1.03)	0.87 (0.78-0.97)	Ref.	0.95 (0.87-1.04)	0.95 (0.87-1.04)
Model 1 [d]	**HR Per 1 SD Increase (95% CI**	0.94 (0.89-0.99)			0.97 (0.92-1.02)		
	P-trend	0.045			0.195		
	HR (95% CI)	Ref.	0.93 (0.83-1.04)	0.86 (0.74-0.98)	Ref.	0.96 (0.85-1.08)	0.95 (0.86-1.04)
Model 2 [d, e]	**HR Per 1 SD Increase (95% CI**	0.90 (0.82-0.98)			0.96 (0.91-1.01)		
	P-trend	0.021			0.129		

Table 6-4_continued *Risk of Bladder Cancer According to Intake of Total Dietary Fibre and Individual Source of Dietary Fibre*

Subgroup		Fruit Fibre (g/day)			Vegetable Fibre (g/day)		
Intake Tertiles		Low (0-2)	Medium (2-4)	High (>4)	Low (0-5)	Medium (5-9)	High (>9)
No. Case/ Participants		1,059/191,576	950/191,613	1,205/191,537	1,185/191,576	1,223/191,575	806/191,575
	HR (95% CI)	Ref.	0.97 (0.88-1.06)	0.96 (0.88-1.06)	Ref.	0.99 (0.90-1.08)	0.92 (0.82-1.02)
Model 1 [d]	**HR Per 1 SD Increase (95% CI**	0.98 (0.95-1.02)			0.99 (0.97-1.03)		
	P-trend	0.396			0.827		
	HR (95% CI)	Ref.	0.98 (0.87-1.11)	0.97 (0.89-1.07)	Ref.	0.98 (0.89-1.08)	0.91 (0.80-1.05)
Model 2 [d, e]	**HR Per 1 SD Increase (95% CI**	0.98 (0.90-1.06)			0.93 (0.86-1.00)		
	P-trend	0.573			0.042		

[d] Model 1: Adjusted for age (years, continuous), sex (male or female), smoking (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), and total energy intake (kcal/day, continuous).

[d, e] Model 2: Additionally, ethnicity (Caucasian or non-Caucasian, if applicable) , alcohol intake (ml/day, continuous), fruit intake (g/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), and vegetable intake (g/day, continuous).

Abbreviation: HR, hazard ratio; CI, confidence interval; SD, standard deviation; g, gram; kcal, kilocalorie; ml, millilitre.

The intervals of tertiles were categorized as open lower values and ended upper values.

Reference group was lowest intake (tertile 1).

P-trend <0.05 was considered statistically significant.

Table 6-5 *Joint Association of Intake of Total Whole Grain and Total Dietary Fibre with Bladder Cancer*

Total Dietary Fibre (g/day)			Low (0-15)		Medium (15-23)		High (>23)	
			No. Case	HR (95% CI)	No. Case	HR (95% CI)	No. Case	HR (95% CI)
Total Whole Grain (g/day)	Model 1 [d]	Low (0-3)	348	Ref.	93	0.92 (0.73-1.16)	84	0.81 (0.64-1.03)
		Medium (3-8)	421	0.92 (0.80-1.07)	157	0.88 (0.73-1.07)	152	0.70 (0.58-0.85)
		High (>8)	222	0.88 (0.74-1.05)	103	0.76 (0.60-0.96)	153	0.69 (0.57-0.86)
	P-trend		0.023					
	Model 2 [d, e]	Low (0-3)	348	Ref.	93	0.94 (0.73-1.21)	84	0.79 (0.60-1.03)
		Medium (3-8)	421	0.91 (0.79-1.05)	157	0.86 (0.69-1.08)	152	0.71 (0.57-0.89)
		High (>8)	222	0.89 (0.70-1.14)	103	0.82] (0.63-1.09)	153	0.73 (0.56-0.95)
	P-trend		0.034					

[d] Model 1: Adjusted for age (years, continuous), sex (male or female), smoking (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), and total energy intake (kcal/day, continuous).
[d, e] Model 2: Additionally, ethnicity (Caucasian or non-Caucasian, if applicable) , alcohol intake (ml/day, continuous), fruit intake (g/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), and vegetable intake (g/day, continuous).
Abbreviation: HR, hazard ratio; CI, confidence interval; g, gram; kcal, kilocalorie; ml, millilitre.
The intervals of tertiles were categorized as open lower values and ended upper values.
Reference group was lowest intake of both total whole grain and total dietary fibre.
P-trend <0.05 was considered statistically significant.

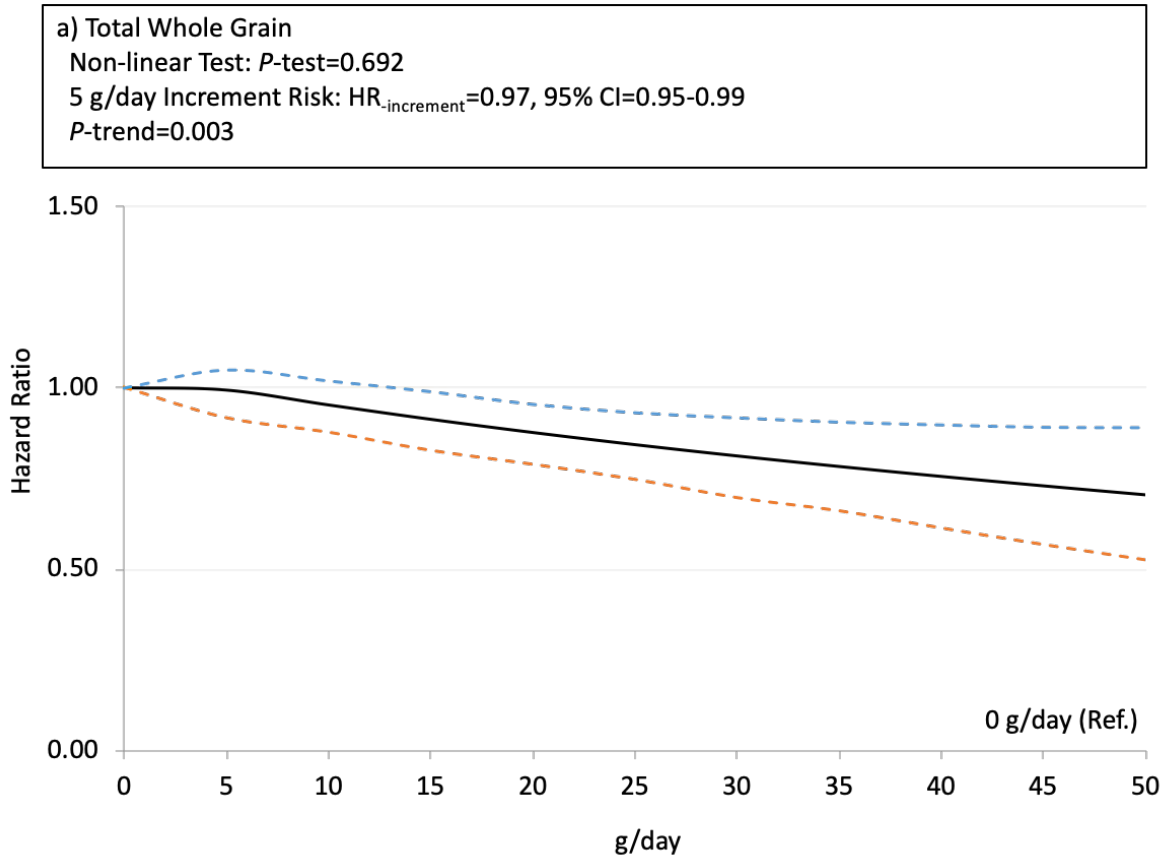
Dose-response Analyses

Dose-response relationships between grains/dietary fibre intake and the risk of bladder cancer are showed in **Figure 6-1** and **Supplementary Table 6-2**. There were inverse associations of intake of total whole grain and total dietary fibre with bladder cancer risk, but no association was observed of intake of total refined grain and cereal fibre with bladder cancer risk. With a significant reduction in risk more than 15 g/day intake of total whole grain and more than 25 g/day intake of total dietary fibre, estimated HRs of 0.97 (0.95-0.99) and 0.96 (0.94-0.98) for per 5 grams daily increment respectively.

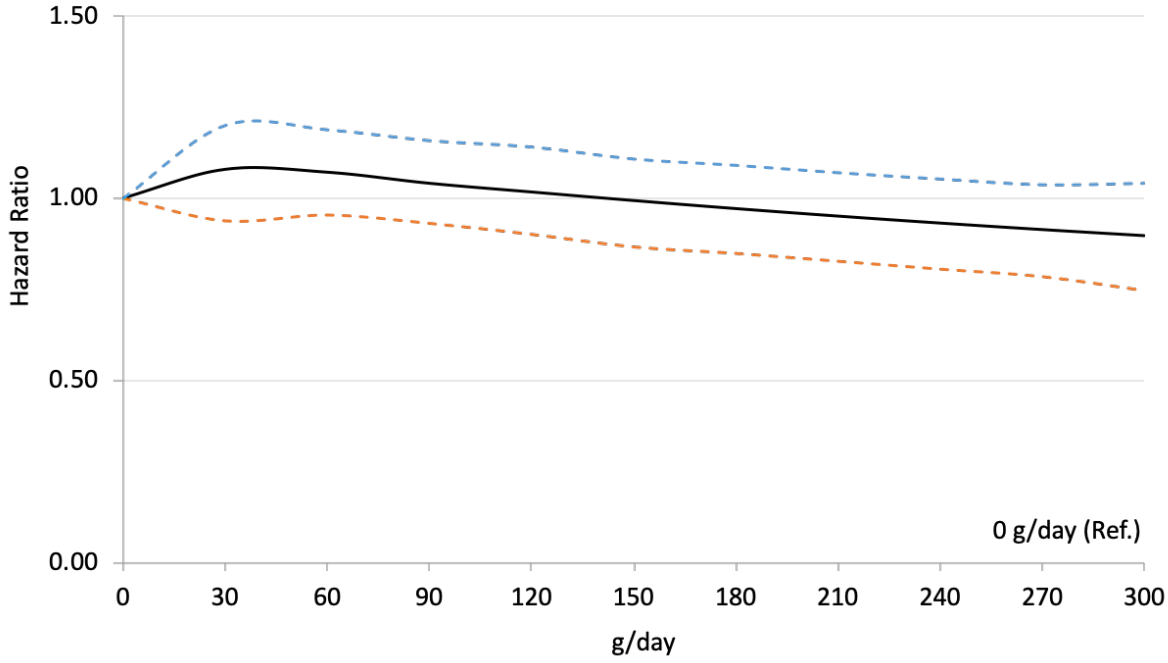
Sensitivity Analyses

Removing bladder cancer cases diagnosed within the first 2 years after enrolling into each individual study gave similar results, in that a decreased bladder cancer risk was observed for total whole grain (comparing the highest with lowest intake tertile: HR_{model2}=0.81, 95% CI=0.70-0.93, P=0.035 for trend; HR per 1 SD increment=0.94, 95% CI=0.90-1.00) and for total dietary fibre (comparing the highest with lowest intake tertile: HR_{model2}=0.85, 95% CI=0.78-0.94, P=0.002 for trend; HR per 1 SD increment=0.86, 95% CI=0.78-0.94) (**Supplementary Table 6-3**). In addition, excluding pasta as a source of total refined grain showed the same results when pasta intake was included (**Supplementary Table 6-4**). The analysis adjusting for

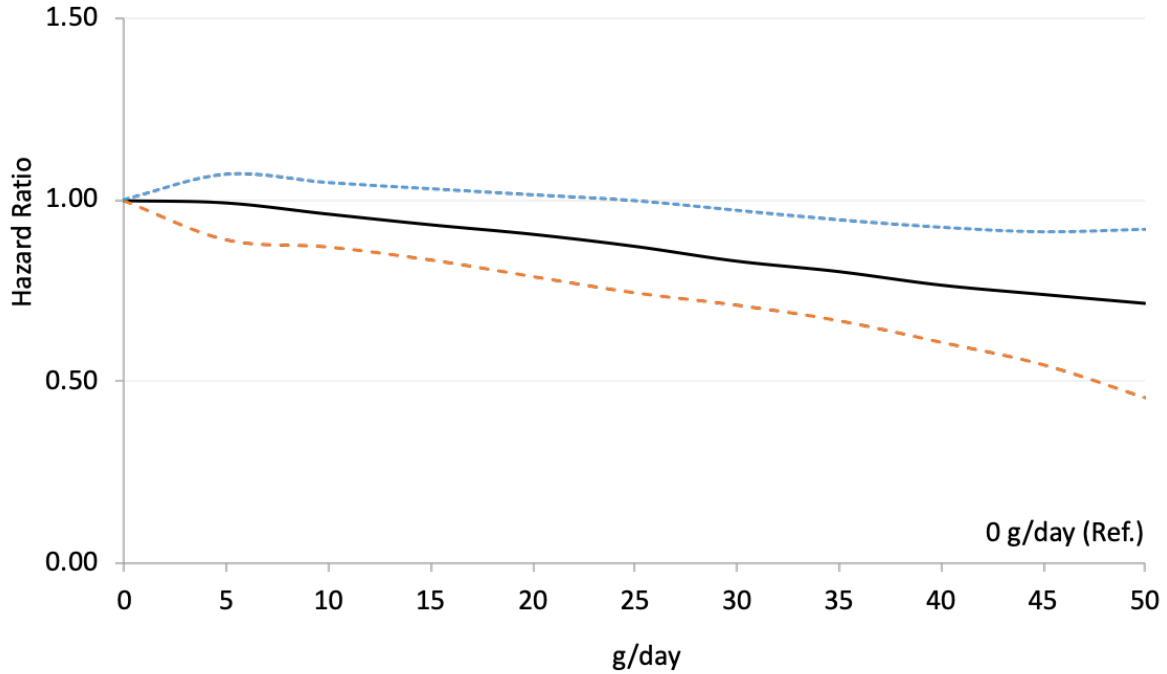
both smoking status and smoking pack-years showed similar results with the adjustment of smoking dummy variable (**Supplementary Table 6-5**). The meta-analyses approach showed similar results, that is a significantly reduced bladder cancer risk with total grain intake 0.89 (95% CI 0.81-0.98) and total dietary fibre intake 0.89 (95% CI 0.80-0.98) and a reduced but non-significant association 0.91 (95% CI 0.19-1.03) for total refined grain intake.



b) Total Refined Grain
Non-linear Test: *P*-test=0.010
30 g/day Increment Risk: $HR_{\text{-increment}}$=1.01, 95% CI=0.98-1.03
P-trend=0.571



c) Total Dietary Fibre
Non-linear Test: *P*-test=0.735
5 g/day Increment Risk: $HR_{\text{-increment}}$=0.96, 95% CI=0.94-0.98
P-trend=0.004



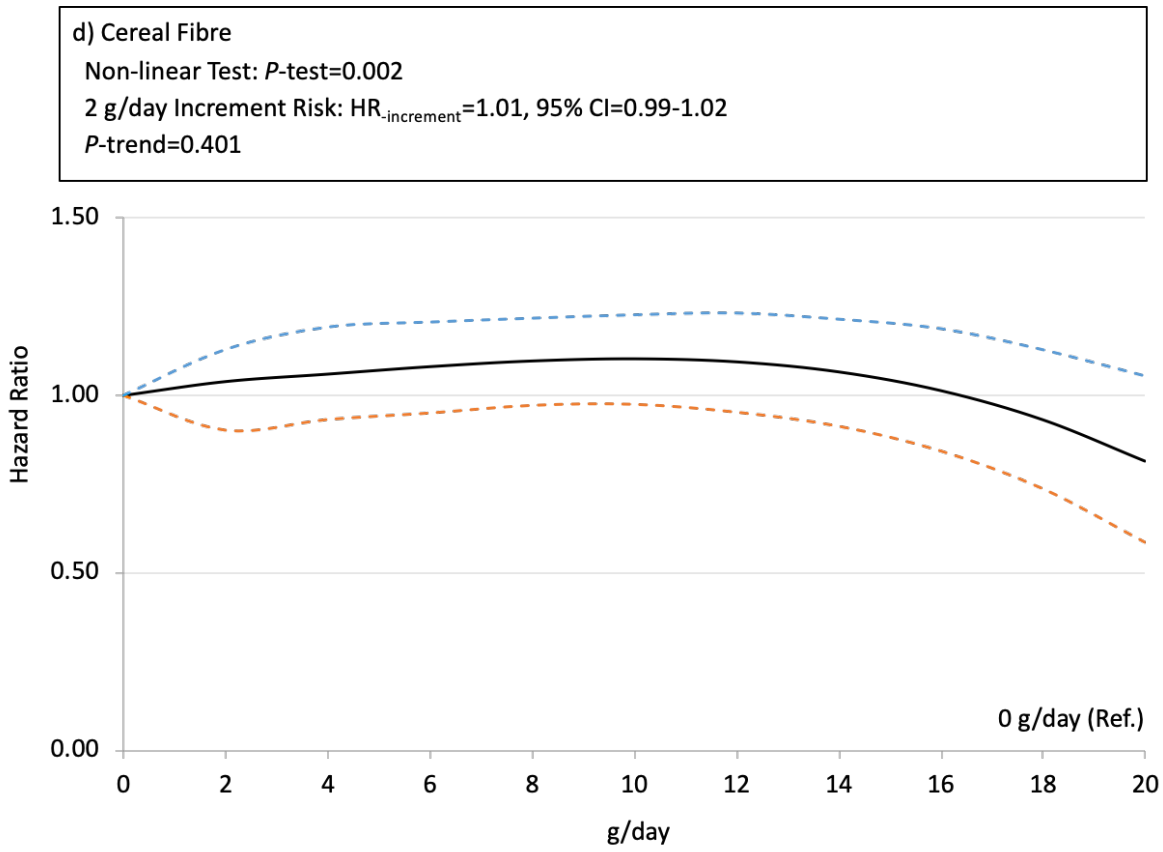


Figure 6-1 *Dose-response Relationships between Grains/Dietary Fibre Intake and the Risk of Bladder Cancer among a) Total Whole Grain; b) Total Refined Grain; c) Total Dietary Fibre; d) Cereal Fibre*

The solid lines represent the hazard ratios (HRs); the dashed lines represent the 95% confidence intervals (CIs) for the trend.

The HRs were adjusted for age (years, continuous), sex (male or female), smoking (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year, and no information on pack-years)]), ethnicity (Caucasian or non-Caucasian), total energy intake (kcal/day, continuous), alcohol intake (ml/day, continuous), sugar intake (g/day, continuous), meat intake (g/day, continuous), vegetable intake (g/day, continuous), fruit intake (g/day, continuous), and fat intake (g/day, continuous) (model 2).

Abbreviation: HR, hazard ratio; CI, confidence interval; g, gram; kcal, kilocalorie; ml, millilitre.

Referent group was non-intake.

P-test <0.05 was considered statistically significant for non-linearity.

P-increment <0.05 was considered statistically significant.

Discussion

This large, multi-centre prospective cohort study indicates a potential beneficial effect of total whole grain and total dietary fibre intake for the prevention of bladder cancer. After controlling for potential confounders, we found that participants with higher intake of total whole grains, particularly brown rice, had a lower bladder cancer risk. Total dietary fibre intake was also inversely associated with bladder cancer risk, whereas intake of individual fibre sources (*i.e.* cereal, fruit and vegetable) did not show any significant association with bladder cancer risk.

To our knowledge, this is the first prospective study to investigate the associations between whole grains intake and bladder cancer risk, in which an inverse association was observed. In line with this result a previously conducted case-control study on total whole grain intake and bladder cancer risk reported a non-significant protective effect of higher intake [22], as well as another case-control study conducted in 1980s reported a modest inverse association between intake of total whole grain and bladder cancer [23]. Since it was reported that an increased bladder cancer risk associated with a high dietary glycaemic load [22], which has been reported to be reduced by the postprandial glucose and insulin responses to whole grains intake [60], the authors suggested that any potential benefit of whole grains intake to bladder cancer risk may act by mitigating the carcinogenic effects of hyperglycaemia and hyperinsulinemia [61-63]. Additionally, our findings are strengthened by experimental studies which show that whole grains may exert their potential anti-tumour activity through the reduction of inflammation [64], which is thought to be related to a decreased bladder cancer risk [65].

One plausible reason why total refined grain did not show to be protective for bladder cancer is that refined grains lack high level of dietary fibre and other bioactive compounds which are rich in whole grains [66, 67]. The lack of dietary fibre could especially be emphasised since the present study shows that both high intake of total whole grain and total dietary fibre is associated with a decreased bladder cancer risk. Epidemiological studies have shown that the intake of whole grains and dietary fibre is associated with a lower risk of chronic diseases [68-72], including various cancer types (*e.g.* hepatocellular carcinoma, oesophageal cancer and pancreatic cancer) [73-75]. In addition, experimental studies have reported that dietary fibre

may protect cancer by improving the insulin sensitivity and metabolic regulation, reducing inflammation, modulating gut microbiota, removal of damaged cells, diluting carcinogens, and decreasing transit time [64, 76-80].

Although most standard pastas are made with refined wheat flour, several whole-wheat pastas are available and worldwide consumed in non-trivial amounts [67, 81]. Considering that several cohorts have been conducted after 2000s, the pasta intake could be a mixture of pasta made of refined and whole grains. Unfortunately, we were unable to distinguish between these types of pasta intake. Nevertheless, we only found a borderline decreased bladder cancer risk for medium intake of “pasta or noodles”. In addition, the results of total refined grains maintained after excluding pasta intake. Since the lack of information on pasta types might have led to inaccurate estimates, results from the present study on the influence of pasta intake could be due to chance, therefore, must be interpreted with caution. Further investigation of the influence of specific pasta types on bladder cancer risk may be warranted.

Half of the world's population relies on rice as a daily staple food and it is considered one of the most important crops globally. The association of this specific grain with bladder cancer risk might differ from any other grains, since rice has been reported to contain considerable amount of inorganic arsenic [82], which has been acknowledged as to be a human carcinogen for bladder [83]. A previous prospective study in the United States found no association for bladder cancer risk with either brown rice or white rice [84]. In our present study we showed no association between white (refined) rice and bladder cancer risk but higher brown (whole grain) rice intake was associated with a decreased bladder cancer risk. It, therefore, remains uncertain whether any beneficial association suggested for whole grains might negate the harmful effect of inorganic arsenic contaminant in rice. Future research should clarify this.

There is convincing evidence that higher intake of dietary fibre is associated with a lower risk of common diseases such as obesity, type 2 diabetes, cardiovascular disease and certain cancer [85]. However, evidence on the influence of dietary fibre on bladder cancer risk is lacking. In the present study we found that high intake of total dietary fibre was significantly associated with a lower bladder cancer risk. However, none of the dietary fibre sources showed to be associated with bladder cancer risk individually. One possible explanation is

that individual dietary fibre intakes failed to reach the threshold at which a protective effect is manifested. In addition, some residual confounding such as “cooking methods” might mask the potential beneficial effect of cereal-, fruit- or vegetable- fibre intake.

Although it is important to evaluate the potential health benefits of whole grains and dietary fibre individually, the present study also investigated the joint effect of those two dietary intakes. Results showed that higher intake of both whole grains and dietary fibres a stronger inverse association with bladder cancer risk than individually intake. This finding furtherly confirmed the assumption that potential benefits of fibre may be from the combination of nutrients of whole grains working together, which suggest the simultaneously intake of both whole grains and dietary fibre as part of a healthy nutritious diet. Although current dietary guidelines recommend individuals at least half of their grain intake to be of whole grains, no quantifications are giving for the amount of total whole grain intake to be consumed [86]. The present study shows that a daily intake of at least 15 grams (uncooked) should be consumed in order to reduce the bladder cancer risk. For dietary fibre, the European Commission’s “health promotion and disease prevention” strategy recommends a daily dietary fibre intake 25 to 38 g in order to prevent non-communicable diseases [87]. This is in accordance with our observation of 25 g/day of total dietary intake for reducing bladder cancer.

For the present study, data were pooled from 13 prospective cohorts, to obtain a sample size, large enough to permit detailed analyses with good precision. The study also has some limitations: a) other than age, sex and smoking, limited information was available on other possible risk factors, for the development of bladder cancer, such as *body mass index* (BMI), physical activity, *socioeconomic status* (SES), disinfection by-products, arsenic in the drinking water, and occupational exposures to potentially carcinogenic chemicals. Although, adjustments for these factors could have influenced the results, the current literature shows that only a small proportion of bladder cancer cases can be attributed to these factors [72]; b) some tumour subtype (*i.e.* muscle-invasive/non-muscle-invasive) information was missing, which hampered the statistical power required for stratified subgroup analyses; c) people with a high intake of whole grains might have generally healthier lifestyles and diets than those with a low intake [88, 89], thus we could not rule out the possibility that some of the associations could be more likely due to a healthy

lifestyle than to whole grains intake per se; d) sample size for whole grains was significantly smaller than for refined grains and dietary fibre, which may have led to less statistical power; e) data was not available on dietary fibre types (*e.g.* soluble versus insoluble), thus we were unable to investigate the association by dietary fibre subtypes; f) although we similar results were found after additional adjusting for potential dietary risk factors, it is still possible that the observed associations were confounded by other dietary constituents or additives (*e.g.* dip or sauce); g) for most cohorts, the exposure variable was assessed by FFQs. Therefore, measurement error and misclassification of study participants in terms of the exposure and outcome are unavoidable. The issue encountered in trying to measure grains intake also highlighted the need for future studies to identify biomarkers for grains intake which would considerably reduce the likelihood of measurement errors and enhance the accuracy of the risk estimates; h) the present study sample consists mostly of Caucasians, and this may limit the generalizability of our results to other racial/ethnic populations or geographic regions; i) there is limited mechanistic evidence to interpret our findings, therefore, further studies are needed to clarify whether there is an underlying mechanism for the association between whole grains /dietary fibre intake and bladder cancer risk.

Conclusion

In summary, our study, of over 3,200 cases of incident bladder cancer occurring in almost 0.6 million participants, indicates a significant inverse association of total whole grain and total dietary fibre intake with bladder cancer risk individually and jointly. This, in turn, supports recommendations to increase the intake of whole grains and dietary fibre to improve public health.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69(1):7-34. doi:10.3322/caac.21551.

2. Richters A, Aben KKH, Kiemeney L. The global burden of urinary bladder cancer: an update. World J Urol. 2019. doi:10.1007/s00345-019-02984-4.

3. Antoni S, Ferlay J, Soerjomataram I, Znaor A, Jemal A, Bray F. Bladder Cancer Incidence and Mortality: A Global Overview and Recent Trends. *Eur Urol* 2017;71(1):96-108. doi:10.1016/j.eururo.2016.06.010.

4. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011;61(2):69-90. doi:10.3322/caac.20107.

5. Ploeg M, Aben KK, Kiemeney LA. The present and future burden of urinary bladder cancer in the world. *World J Urol* 2009;27(3):289-293. doi:10.1007/s00345-009-0383-3.

6. Marugame T, Mizuno S. Comparison of prostate cancer mortality in five countries: France, Italy, Japan, UK and USA from the WHO mortality database (1960–2000). *Jpn J Clin Oncol* 2005;35(11):690-691. doi:10.1093/jjco/hyi185.

7. Stewart B, Wild CP. World cancer report 2014. *International Agency for Research on Cancer, World Health Organization*: WHO Press, 2017.

8. Letašiová S, Medveďová A, Šovčíková A *et al*. Bladder cancer, a review of the environmental risk factors. *Environmental Health* 2012;11(Suppl 1):S11. doi:10.1186/1476-069X-11-S1-S11.

9. Botteman MF, Pashos CL, Redaelli A, Laskin B, Hauser R. The health economics of bladder cancer. *Pharmacoeconomics* 2003;21(18):1315-1330. doi:10.1007/BF03262330.

10. Al-Zalabani AH, Stewart KF, Wesselius A, Schols AM, Zeegers MP. Modifiable risk factors for the prevention of bladder cancer: a systematic review of meta-analyses. European journal of epidemiology. 2016;31(9):811-51. doi:10.1007/s10654-016-0138-6.

11. Piyathilake C. Dietary factors associated with bladder cancer. Investigative and clinical urology. 2016;57 Suppl 1:S14-25. doi:10.4111/icu.2016.57.S1.S14.

12. Johansson SL, Cohen SM. Epidemiology and etiology of bladder cancer. Semin Surg Oncol. 1997;13(5):291-8. doi:10.1002/(sici)1098-2388(199709/10)13:5<291::aid-ssu2>3.0.co;2-8.

13. Thun M, Linet MS, Cerhan JR, Haiman CA, Schottenfeld D. Cancer epidemiology and prevention. Oxford University Press; 2017.

14. Oberoi S, Barchowsky A, Wu F. The global burden of disease for skin, lung, and bladder cancer caused by arsenic in food. Cancer epidemiology, biomarkers & prevention : a

publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2014;23(7):1187-94. doi:10.1158/1055-9965.Epi-13-1317.

15. European Food Information Council (EUFIC). https://wwweuficorg/en.Access July 15, 2019.

16. US Department of Agriculture and US Department of Health and Human Services. Dietary guidelines for Americans. https://wwwchoosemyplategov/dietary-guidelines.Accessed July 15, 2019.

17. Mozaffarian RS, Lee RM, Kennedy MA, Ludwig DS, Mozaffarian D, Gortmaker SL. Identifying whole grain foods: a comparison of different approaches for selecting more healthful whole grain products. Public health nutrition. 2013;16(12):2255-64. doi:10.1017/s1368980012005447.

18. U.S. Department of Health and Human Services and U.S. Department of Agriculture. 2015 – 2020 Dietary Guidelines for Americans. 8th Edition. December 2015. Available at http://healthgov/dietaryguidelines/2015/guidelines/.Accessed September 7, 2019.

19. Makarem N, Nicholson JM, Bandera EV, McKeown NM, Parekh N. Consumption of whole grains and cereal fiber in relation to cancer risk: a systematic review of longitudinal studies. Nutr Rev. 2016;74(6):353-73. doi:10.1093/nutrit/nuw003.

20. Hsu IR, Kim SP, Kabir M, Bergman RN. Metabolic syndrome, hyperinsulinemia, and cancer. The American journal of clinical nutrition. 2007;86(3):s867-71. doi:10.1093/ajcn/86.3.867S.

21. Murata M. Inflammation and cancer. Environ Health Prev Med. 2018;23(1):50. doi:10.1186/s12199-018-0740-1.

22. Augustin LSA, Taborelli M, Montella M, Libra M, La Vecchia C, Tavani A et al. Associations of dietary carbohydrates, glycaemic index and glycaemic load with risk of bladder cancer: a case-control study. The British journal of nutrition. 2017;118(9):722-9. doi:10.1017/s0007114517002574.

23. La Vecchia C, Chatenoud L, Negri E, Franceschi S. Session: whole cereal grains, fibre and human cancer wholegrain cereals and cancer in Italy. The Proceedings of the Nutrition Society. 2003;62(1):45-9. doi:10.1079/pns2002235.

24. Nicodemus KK, Jacobs DR, Jr., Folsom AR. Whole and refined grain intake and risk of incident postmenopausal breast cancer (United States). Cancer causes & control : CCC. 2001;12(10):917-25.

25. Farvid MS, Cho E, Eliassen AH, Chen WY, Willett WC. Lifetime grain consumption and breast cancer risk. Breast cancer research and treatment. 2016;159(2):335-45. doi:10.1007/s10549-016-3910-0.

26. Kasum CM, Jacobs DR, Jr., Nicodemus K, Folsom AR. Dietary risk factors for upper aerodigestive tract cancers. International journal of cancer. 2002;99(2):267-72. doi:10.1002/ijc.10341.

27. Kasum CM, Nicodemus K, Harnack LJ, Jacobs DR, Jr., Folsom AR. Whole grain intake and incident endometrial cancer: the Iowa Women's Health Study. Nutrition and cancer. 2001;39(2):180-6. doi:10.1207/S15327914nc392_4.

28. Makarem N, Bandera EV, Lin Y, McKeown NM, Hayes RB, Parekh N. Associations of Whole and Refined Grain Intakes with Adiposity-Related Cancer Risk in the Framingham Offspring Cohort (1991-2013). Nutrition and cancer. 2018;70(5):776-86. doi:10.1080/01635581.2018.1470647.

29. Xu Y, Yang J, Du L, Li K, Zhou Y. Association of whole grain, refined grain, and cereal consumption with gastric cancer risk: A meta-analysis of observational studies. Food Sci Nutr. 2019;7(1):256-65. doi:10.1002/fsn3.878.

30. Schwingshackl L, Schwedhelm C, Hoffmann G, Knuppel S, Laure Preterre A, Iqbal K et al. Food groups and risk of colorectal cancer. International journal of cancer. 2018;142(9):1748-58. doi:10.1002/ijc.31198.

31. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol. 1997;26 Suppl 1:S6-14. doi:10.1093/ije/26.suppl_1.s6.

32. Tjonneland A, Olsen A, Boll K, Stripp C, Christensen J, Engholm G et al. Study design, exposure variables, and socioeconomic determinants of participation in Diet, Cancer and Health: a population-based prospective cohort study of 57,053 men and women in Denmark. Scandinavian journal of public health. 2007;35(4):432-41. doi:10.1080/14034940601047986.

33. Clavel-Chapelon F, van Liere MJ, Giubout C, Niravong MY, Goulard H, Le Corre C et al. E3N, a French cohort study on cancer risk factors. E3N Group. Etude Epidemiologique aupres de femmes de l'Education Nationale. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP). 1997;6(5):473-8. doi:10.1097/00008469-199710000-00007.

34. Boeing H, Korfmann A, Bergmann MM. Recruitment procedures of EPIC-Germany. European Investigation into Cancer and Nutrition. Annals of nutrition & metabolism. 1999;43(4):205-15. doi:10.1159/000012787.

35. 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. doi:10.1079/phn2002394.

36. Panico S, Dello Iacovo R, Celentano E, Galasso R, Muti P, Salvatore M et al. Progetto ATENA, a study on the etiology of major chronic diseases in women: design, rationale and objectives. European journal of epidemiology. 1992;8(4):601-8.

37. Manjer J, Carlsson S, Elmstahl S, Gullberg B, Janzon L, Lindstrom M et al. The Malmo Diet and Cancer Study: representativity, cancer incidence and mortality in participants and non-participants. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP). 2001;10(6):489-99. doi:10.1097/00008469-200112000-00003.

38. 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. doi:10.1080/14034950310001432.

39. Beulens JW, Monninkhof EM, Verschuren WM, van der Schouw YT, Smit J, Ocke MC et al. Cohort profile: the EPIC-NL study. Int J Epidemiol. 2010;39(5):1170-8. doi:10.1093/ije/dyp217.

40. Davey GK, Spencer EA, Appleby PN, Allen NE, Knox KH, Key TJ. EPIC-Oxford: lifestyle characteristics and nutrient intakes in a cohort of 33 883 meat-eaters and 31 546 non meat-eaters in the UK. Public health nutrition. 2003;6(3):259-69. doi:10.1079/phn2002430.

41. Day N, Oakes S, Luben R, Khaw KT, Bingham S, Welch A et al. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. British journal of cancer. 1999;80 Suppl 1:95-103.

42. Lund E, Dumeaux V, Braaten T, Hjartaker A, Engeset D, Skeie G et al. Cohort profile: The Norwegian Women and Cancer Study--NOWAC--Kvinner og kreft. Int J Epidemiol. 2008;37(1):36-41. doi:10.1093/ije/dym137.

43. van den Brandt PA, Goldbohm RA, van 't Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. Journal of clinical epidemiology. 1990;43(3):285-95. doi:10.1016/0895-4356(90)90009-E.

44. White E, Patterson RE, Kristal AR, Thornquist M, King I, Shattuck AL et al. VITamins And Lifestyle cohort study: study design and characteristics of supplement users. Am J Epidemiol. 2004;159(1):83-93. doi:10.1093/aje/kwh010.

45. Giles GG, English DR. The Melbourne Collaborative Cohort Study. IARC Sci Publ. 2002;156:69-70.

46. Milne RL, Fletcher AS, MacInnis RJ, Hodge AM, Hopkins AH, Bassett JK et al. Cohort Profile: The Melbourne Collaborative Cohort Study (Health 2020). Int J Epidemiol. 2017;46(6):1757-i. doi:10.1093/ije/dyx085.

47. Goossens ME, Isa F, Brinkman M *et al*. International pooled study on diet and bladder cancer: the bladder cancer, epidemiology and nutritional determinants (BLEND) study: design and baseline characteristics. *Archives of Public Health* 2016;74(1):30-39. doi:10.1186/s13690-016-0140-1.

48. Bassett JK, English DR, Fahey MT, Forbes AB, Gurrin LC, Simpson JA et al. Validity and calibration of the FFQ used in the Melbourne Collaborative Cohort Study. Public health nutrition. 2016;19(13):2357-68. doi:10.1017/s1368980016000690.

49. Hodge A, Patterson AJ, Brown WJ, Ireland P, Giles G. The Anti Cancer Council of Victoria FFQ: relative validity of nutrient intakes compared with weighed food records in young to middle-aged women in a study of iron supplementation. Aust N Z J Public Health. 2000;24(6):576-83. doi:10.1111/j.1467-842x.2000.tb00520.x.

50. Kaaks R, Riboli E. Validation and calibration of dietary intake measurements in the EPIC project: methodological considerations. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol. 1997;26 Suppl 1:S15-25. doi:10.1093/ije/26.suppl_1.s15.

51. Zeegers MP, Goldbohm RA, van den Brandt PA. Are retinol, vitamin C, vitamin E, folate and carotenoids intake associated with bladder cancer risk? Results from the Netherlands Cohort Study. British journal of cancer. 2001;85(7):977-83. doi:10.1054/bjoc.2001.1968.

52. Ferrari P, Slimani N, Ciampi A, Trichopoulou A, Naska A, Lauria C et al. Evaluation of under- and overreporting of energy intake in the 24-hour diet recalls in the European Prospective Investigation into Cancer and Nutrition (EPIC). Public health nutrition. 2002;5(6b):1329-45. doi:10.1079/phn2002409.

53. Poortvliet E, Klensin J, Kohlmeier L. Rationale document for the Eurocode 2 food coding system (version 91/2). *Eur J Clin Nutr* 1992;46(Suppl 5):S9-S24.

54. Watt BK, Merrill AL, Orr ML, Wu W, Pecot R. Composition of foods-raw, processed, prepared. Washington: Consumer and Food Economics Institute, Agricultural Research Service, US Department of Agriculture. 1964.

55. Giganti MJ, Luz PM, Caro-Vega Y, Cesar C, Padgett D, Koenig S et al. A Comparison of Seven Cox Regression-Based Models to Account for Heterogeneity Across Multiple HIV Treatment Cohorts in Latin America and the Caribbean. AIDS research and human retroviruses. 2015;31(5):496-503. doi:10.1089/aid.2014.0241.

56. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. The American journal of clinical nutrition. 1997;65(4 Suppl):1220S-8S; discussion 9S-31S. doi:10.1093/ajcn/65.4.1220S.

57. Bagnardi V, Zambon A, Quatto P, Corrao G. Flexible meta-regression functions for modeling aggregate dose-response data, with an application to alcohol and mortality. American journal of epidemiology. 2004;159(11):1077-86. doi:https://doi.org/10.1093/aje/kwh142.

58. Royston P, Altman DG. Approximating statistical functions by using fractional polynomial regression. Journal of the Royal Statistical Society: Series D (The Statistician). 1997;46(3):411-22. doi:https://doi.org/10.1111/1467-9884.00093.

59. Jones BL, Nagin DS. A Stata plugin for estimating group-based trajectory models. Research Showcase@ CMU Carnegie Mellon University Retrieved on July. 2012;10:2015. doi:10.1093/aje/kwh142.

60. Esmaillzadeh A, Mirmiran P, Azizi F. Whole-grain consumption and the metabolic syndrome: a favorable association in Tehranian adults. Eur J Clin Nutr. 2005;59(3):353-62. doi:10.1038/sj.ejcn.1602080.

61. Jee SH, Ohrr H, Sull JW, Yun JE, Ji M, Samet JM. Fasting serum glucose level and cancer risk in Korean men and women. Jama. 2005;293(2):194-202. doi:10.1001/jama.293.2.194.

62. Liu S, Li Y, Lin T, Fan X, Liang Y, Heemann U. High dose human insulin and insulin glargine promote T24 bladder cancer cell proliferation via PI3K-independent activation of Akt. Diabetes Res Clin Pract. 2011;91(2):177-82. doi:10.1016/j.diabres.2010.11.009.

63. Ornskov D, Nexo E, Sorensen BS. Insulin-induced proliferation of bladder cancer cells is mediated through activation of the epidermal growth factor system. Febs j. 2006;273(23):5479-89. doi:10.1111/j.1742-4658.2006.05539.x.

64. Qi L, van Dam RM, Liu S, Franz M, Mantzoros C, Hu FB. Whole-grain, bran, and cereal fiber intakes and markers of systemic inflammation in diabetic women. Diabetes care. 2006;29(2):207-11. doi:10.2337/diacare.29.02.06.dc05-1903.

65. Gakis G. The role of inflammation in bladder cancer. Advances in experimental medicine and biology. 2014;816:183-96. doi:10.1007/978-3-0348-0837-8_8.

66. Slavin J. Why whole grains are protective: biological mechanisms. The Proceedings of the Nutrition Society. 2003;62(1):129-34. doi:10.1079/pns2002221.

67. Frolich W, Aman P, Tetens I. Whole grain foods and health - a Scandinavian perspective. Food Nutr Res. 2013;57. doi:10.3402/fnr.v57i0.18503.

68. de Munter JS, Hu FB, Spiegelman D, Franz M, van Dam RM. Whole grain, bran, and germ intake and risk of type 2 diabetes: a prospective cohort study and systematic review. PLoS medicine. 2007;4(8):e261. doi:10.1371/journal.pmed.0040261.

69. Ye EQ, Chacko SA, Chou EL, Kugizaki M, Liu S. Greater whole-grain intake is associated with lower risk of type 2 diabetes, cardiovascular disease, and weight gain. The Journal of nutrition. 2012;142(7):1304-13. doi:10.3945/jn.111.155325.

70. Cho SS, Qi L, Fahey GC, Jr., Klurfeld DM. Consumption of cereal fiber, mixtures of whole grains and bran, and whole grains and risk reduction in type 2 diabetes, obesity, and cardiovascular disease. The American journal of clinical nutrition. 2013;98(2):594-619. doi:10.3945/ajcn.113.067629.

71. Steffen LM, Jacobs DR, Jr., Murtaugh MA, Moran A, Steinberger J, Hong CP et al. Whole grain intake is associated with lower body mass and greater insulin sensitivity among adolescents. Am J Epidemiol. 2003;158(3):243-50. doi:10.1093/aje/kwg146.

72. Mellen PB, Walsh TF, Herrington DM. Whole grain intake and cardiovascular disease: a meta-analysis. Nutrition, metabolism, and cardiovascular diseases : NMCD. 2008;18(4):283-90. doi:10.1016/j.numecd.2006.12.008.

73. Yang W, Ma Y, Liu Y, Smith-Warner SA, Simon TG, Chong DQ et al. Association of Intake of Whole Grains and Dietary Fiber With Risk of Hepatocellular Carcinoma in US Adults. JAMA Oncol. 2019;5(6):879-86. doi:10.1001/jamaoncol.2018.7159.

74. Skeie G, Braaten T, Olsen A, Kyro C, Tjonneland A, Landberg R et al. Intake of whole grains and incidence of oesophageal cancer in the HELGA Cohort. European journal of epidemiology. 2016;31(4):405-14. doi:10.1007/s10654-015-0057-y.

75. Lei Q, Zheng H, Bi J, Wang X, Jiang T, Gao X et al. Whole Grain Intake Reduces Pancreatic Cancer Risk: A Meta-Analysis of Observational Studies. Medicine (Baltimore). 2016;95(9):e2747. doi:10.1097/md.0000000000002747.

76. Te Morenga L, Docherty P, Williams S, Mann J. The Effect of a Diet Moderately High in Protein and Fiber on Insulin Sensitivity Measured Using the Dynamic Insulin Sensitivity and Secretion Test (DISST). Nutrients. 2017;9(12). doi:10.3390/nu9121291.

77. Kasubuchi M, Hasegawa S, Hiramatsu T, Ichimura A, Kimura I. Dietary gut microbial metabolites, short-chain fatty acids, and host metabolic regulation. Nutrients. 2015;7(4):2839-49. doi:10.3390/nu7042839.

78. Hemarajata P, Versalovic J. Effects of probiotics on gut microbiota: mechanisms of intestinal immunomodulation and neuromodulation. Therap Adv Gastroenterol. 2013;6(1):39-51. doi:10.1177/1756283x12459294.

79. Aron-Wisnewsky J, Clement K. The gut microbiome, diet, and links to cardiometabolic and chronic disorders. Nat Rev Nephrol. 2016;12(3):169-81. doi:10.1038/nrneph.2015.191.

80. Aune D, Keum N, Giovannucci E, Fadnes LT, Boffetta P, Greenwood DC et al. Whole grain consumption and risk of cardiovascular disease, cancer, and all cause and cause specific mortality: systematic review and dose-response meta-analysis of prospective studies. BMJ (Clinical research ed). 2016;353:i2716. doi:10.1136/bmj.i2716.

81. Chatenoud L, Tavani A, La Vecchia C, Jacobs DR, Jr., Negri E, Levi F et al. Whole grain food intake and cancer risk. International journal of cancer. 1998;77(1):24-8. doi:10.1002/(sici)1097-0215(19980703)77:1<24::aid-ijc5>3.0.co;2-1.

82. Kumarathilaka P, Seneweera S, Ok YS, Meharg A, Bundschuh J. Arsenic in cooked rice foods: Assessing health risks and mitigation options. Environment international. 2019;127:584-91. doi:10.1016/j.envint.2019.04.004.

83. ARC. Arsenic, metals, fibers and dusts. A review of human carcinogens. IARC monographs on the evaluation of carcinogenic risks to humans. 2012;100C. Available at: https://monographsiarcfr/wp-content/uploads/2018/06/ Accessed September 17, 2019.

84. Zhang R, Zhang X, Wu K, Wu H, Sun Q, Hu FB et al. Rice consumption and cancer incidence in US men and women. International journal of cancer. 2016;138(3):555-64. doi:10.1002/ijc.29704.

85. Anderson JW, Baird P, Davis RH, Jr., Ferreri S, Knudtson M, Koraym A et al. Health benefits of dietary fiber. Nutrition reviews. 2009;67(4):188-205. doi:10.1111/j.1753-4887.2009.00189.x.

86. Marmot M, Atinmo T, Byers T, Chen J, Hirohata T, Jackson A et al. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. 2007.

87. The European Commission's science and knowledge service. Available at https://eceuropaeu/jrc/en/health-knowledge-gateway/promotion-prevention/nutrition/fibre.Accessed September 17, 2019.

88. Kyro C, Skeie G, Dragsted LO, Christensen J, Overvad K, Hallmans G et al. Intake of whole grains in Scandinavia is associated with healthy lifestyle, socio-economic and dietary factors. Public health nutrition. 2011;14(10):1787-95. doi:10.1017/s1368980011000206.

89. Wu H, Flint AJ, Qi Q, van Dam RM, Sampson LA, Rimm EB et al. Association between dietary whole grain intake and risk of mortality: two large prospective studies in US men and women. JAMA internal medicine. 2015;175(3):373-84. doi:10.1001/jamainternmed.2014.6283.

Supplementary Materials to Chapter 6

Supplementary Table 6-1 Additional Baseline Characteristic of the Participant Cohort Studies

Supplementary Table 6-2 Adjusted Hazard Ratios and 95% Confidence Intervals of Each Dose-response Relationship between Grains/Dietary Fibre Intake and the Risk of Bladder Cancer

Supplementary Table 6-3 Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Grains and Dietary Fibre Intake (After removing 358 bladder cancer cases diagnosed within 2 years)

Supplementary Table 6-4 Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Refined Grain Products Intake (After Removing Pasta Intake)

Supplementary Table 6-5 Adjusted Hazard Ratios and 95% Confidence Intervals of Bladder Cancer According to Intake of Total Whole Grain, Total Refined Grain and Total Dietary Fibre in Different Adjustment Models

Supplementary Figure 6-1 Forest Plot of Meta-Analyses with HRs and 95% CIs for High vs. Low Intake of Total Whole Grain, Total Refined Grain and Total Dietary Fibre with Bladder Cancer Risk

Supplementary Table 6-1 *Additional Baseline Characteristic of Included Studies*

Cohort	No. Participants (%) [a]	Follow-up Years [b]	Age at Baseline [c]	Person years	No. Incidences	Years to Diagnosis [d]	% of Ever Smokers	Initiate year
EPIC (Europe)								
Denmark	56,005	10.93	50-66	612,129	391	6.82	0.68	1993
France	64,866	10.39	42-71	673,878	31	7.25	0.52	1993
Germany	49,457	9.88	19-70	488,778	207	6.23	0.40	1994
Greece	25,268	9.51	19-84	240,400	50	5.59	0.65	1993
Italy	45,204	11.23	24-78	507,763	187	6.89	0.29	1992
Spain	40,782	12.06	29-70	491,927	152	7.75	0.54	1992
Sweden	49,326	13.12	29-74	647,121	303	8.24	0.44	1991
Netherlands	37,102	11.81	20-70	438,312	107	7.36	0.55	1993
United Kingdom	75,033	11.14	20-98	835,558	248	7.05	0.45	1993
Norway	33,855	9.73	41-56	329,425	24	5.81	0.51	1998
NLCS (Netherlands)	5,247	14.06	55-70	73,771	877	9.34	0.62	1986
VITAL (USA)	69,491	6.75	50-76	468,779	346	3.87	0.44	2000
MCCS (Australia)	23,090	22.86	35-76	527,859	291	14.74	0.64	990
Overall	574,726	11.02	19-99	6,335,667	3,214	8.06	0.50	1960-2000

Abbreviations: EPIC, European Prospective Investigation into Cancer and Nutrition; NLCS, The Netherlands Cohort Study; VITAL, VITamins And Lifestyle study; MCCS, Melbourne Collaborative Cohort Study.

[a] As a result of exclusion criteria, the cohort sizes and number of cases included in the pooled analyses may differ from original study-specific publications.

[b] Median years of the time interval from two years after study entry to the date of the last follow-up (e.g. date of death, lost to follow-up, or study exit, whichever came first).

[c] Age ranges at baseline (minimum to maximum).

[d] Time-to-bladder cancer defined by median years of study entry to the date of diagnosis among bladder cancer cases.

***Supplementary Table 6-2** Adjusted Hazard Ratios and 95% Confidence Intervals of Each Dose-response Relationship between Grains/Dietary Fibre Intake and the Risk of Bladder Cancer*

Total Whole Grain		Total Refined Grain		Total Dietary Fibre		Cereal Fibre	
Dose (g/day)	HR (95% CI)	Dose (g/day)	HR (95% CI)	Dose (g/day)	HR (95% CI)	Dose (g/day)	HR (95% CI)
0	Ref.	0	Ref.	0	Ref.	0	Ref.
0-5	0.99 (0.92-1.05)	0-30	1.08 (0.94-1.20)	0-5	0.99 (0.89-1.07)	0-2	1.04 (0.90-1.13)
5-10	0.95 (0.92-1.05)	30-60	1.07 (0.95-1.19)	5-10	0.96 (0.87-1.05)	2-4	1.06 (0.93-1.19)
10-15	0.91 (0.88-1.02)	60-90	1.04 (0.93-1.16)	10-15	0.93 (0.84-1.03)	4-6	1.08 (0.95-1.20)
15-20	0.88 (0.83-0.99)	90-120	1.02 (0.90-1.14)	15-20	0.91 (0.79-1.01)	6-8	1.10 (0.97-1.21)
15-25	0.84 (0.75-0.93)	120-150	0.99 (0.87-1.11)	15-25	0.87 (0.75-1.00)	8-10	1.11 (0.99-1.23)
25-30	0.81 (0.70-0.92)	150-180	0.97 (0.85-1.09)	25-30	0.83 (0.71-0.97)	10-12	1.10 (0.95-1.23)
30-35	0.78 (0.66-0.90)	180-210	0.95 (0.83-1.07)	30-35	0.80 (0.67-0.94)	12-14	1.07 (0.91-1.21)
35-40	0.76 (0.62-0.90)	210-240	0.93 (0.81-1.05)	35-40	0.77 (0.61-0.92)	14-16	1.01 (0.84-1.18)
40-45	0.73 (0.57-0.89)	240-270	0.91 (0.79-1.04)	40-45	0.74 (0.55-0.91)	16-18	0.93 (0.74-1.13)

Supplementary Table 6-2_continued *Adjusted Hazard Ratios and 95% Confidence Intervals of Each Dose-response Relationship between Grains/Dietary Fibre Intake and the Risk of Bladder Cancer*

Total Whole Grain		Total Refined Grain		Total Dietary Fibre		Cereal Fibre	
Dose (g/day)	**HR (95% CI)**	**Dose (g/day)**	**HR (95% CI)**	**Dose (g/day)**	**HR (95% CI)**	**Dose (g/day)**	**HR (95% CI)**
45-50	0.71 (0.53-0.89)	270-300	0.90 (0.76-1.04)	45-50	0.72 (0.46-0.92)	18-20	0.82 (0.59-1.05)

[e] *Model 2: Adjusted for age (years, continuous), sex (male or female), smoking (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), total energy intake (kcal/day, continuous), ethnicity (Caucasian or non-Caucasian, if applicable) , alcohol intake (ml/day, continuous), fruit intake (g/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), and vegetable intake (g/day, continuous).*

Abbreviation: HR, hazard ratio; CI, confidence interval; g, gram; kcal, kilocalorie; ml, millilitre.

The intervals of tertiles were categorized as open lower values and ended upper values. Reference group was non-intake.

Supplementary Table 6-3 *Risk of Bladder Cancer According to Intake of Total Whole Grain, Total Refined Grain, Total Dietary Fibre and Cereal Fibre (After Removing 358 Bladder Cancer Cases Diagnosed within 2 years)*

Subgroup		Total whole grains (g/day)			Total refined grains (g/day)		
Intake Tertiles		Low (0-3)	Medium (3-8)	High (>8)	Low (0-102)	Medium (102-181)	High (>181)
No. Case/ Participants		910/72,740	323/74,255	360/70,421	863/191,435	1,117/191,454	876/191,479
	HR (95% CI)	Ref.	0.98 (0.85-1.12)	0.83 (0.73-0.94)	Ref.	0.94 (0.85-1.04)	0.97 (0.86-1.10)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	0.95 (0.90-0.99)			0.99 (0.94-1.04)		
	P-trend	0.024			0.701		
	HR (95% CI)	Ref.	0.98 (0.85-1.14)	0.81 (0.70-0.93)	Ref.	0.91 (0.81-1.02)	0.93 (0.81-1.07)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	0.94 (0.90-1.00)			0.98 (0.92-1.04)		
	P-trend	0.035			0.522		

Supplementary Table 6-3_continued *Risk of Bladder Cancer According to Intake of Total Whole Grain, Total Refined Grain, Total Dietary Fibre and Cereal Fibre (After Removing 358 Bladder Cancer Cases Diagnosed within 2 years)*

Subgroup		Total dietary fibre (g/day)			Cereal Fibre (g/day)		
Intake Tertiles		Low (0-15)	Medium (15-23)	High (>23)	Low (0-5)	Medium (5-9)	High (>9)
No. Case/ Participants		871/191,432	983/191,461	1,002/191,475	637/191,457	918/191,468	1,301/191,443
	HR (95% CI)	Ref.	0.93 (0.84-1.04)	0.89 (0.80-0.97)	Ref.	1.02 (0.90-1.15)	1.09 (0.96-1.24)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**	0.93 (0.87-0.99)			1.01 (0.96-1.06)		
	P-trend	0.020			0.702		
	HR (95% CI)	Ref.	0.92 (0.83-1.03)	0.85 (0.76-0.96)	Ref.	1.05 (0.92-1.20)	1.10 (0.96-1.27)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**	0.86 (0.78-0.94)			1.00 (0.95-1.05)		
	P-trend	0.002			0.979		

[e] Model 1: Adjusted for age (years, continuous), sex (male or female), smoking (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), and total energy intake (kcal/day, continuous).

[e, f] Model 2: Additionally, ethnicity (Caucasian or non-Caucasian, if applicable) , alcohol intake (ml/day, continuous), fruit intake (g/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), and vegetable intake (g/day, continuous).

Abbreviation: HR, hazard ratio; SD, standard deviation; CI, confidence interval; MIBC, muscle invasive bladder cancer; NMIBC, non-muscle invasive bladder cancer; g, gram; kcal, kilocalorie; ml, millilitre.

The intervals of tertiles were categorized as open lower values and ended upper values.

Reference group was lowest intake (tertile 1).

P-trend <0.05 was considered statistically significant.

Supplementary Table 6-4 *Risk of Bladder Cancer According to Intake of Total Refined Grain (After Removing Pasta Intake)*

Subgroup		Overall			MIBC			NMIBC	
Intake Tertiles		Low (0-92)	Medium (92-172)	High (>172)	Low (0-92)	Medium (92-172)	High (>172)	Low (0-170)	Medium (170-301)
No. Case/ Participants		917/191,576	1,284/191,575	1,013/191,575	233/190,892	379/190,670	161/190,723	395/191,054	503/190,794
	HR (95% CI)	Ref.	0.98 (0.89-1.08)	1.00 (0.89-1.12)	Ref.	0.99 (0.81-1.21)	0.85 (0.66-1.11)	Ref.	0.97 (0.82-1.15)
Model 1 [e]	**HR Per 1 SD Increase**	0.98 (0.93-1.03)			0.93 (0.81-1.06)			0.96 (0.88-1.04)	
	P-trend	0.365			0.273			0.335	
	HR (95% CI)	Ref.	0.96 (0.86-1.08)	0.97 (0.85-1.11)	Ref.	0.96 (0.76-1.22)	0.77 (0.57-1.05)	Ref.	0.97 (0.79-1.18)
Model 2 [e, f]	**HR Per 1 SD Increase**	0.98 (0.92-1.03)			0.92 (0.78-1.07)			0.94 (0.86-1.04)	
	P-trend	0.405			0.348			0.234	

Supplementary Table 6-4_continued *Risk of Bladder Cancer According to Intake of Total Refined Grain (After Removing Pasta Intake)*

Subgroup		NMIBC	Male			Female		
Intake Tertiles		High (>301)	Low (0-98)	Medium (98-199)	High (>199)	Low (0-90)	Medium (90-159)	High (>159)
No. Case/ Participants		370/190,932	712/62,964	1,112/62,944	592/62,954	271/128,622	304/128,621	223/128,621
	HR (95% CI)	0.92 (0.76-1.13)	Ref.	0.92 (0.82-1.03)	0.91 (0.80-1.05)	Ref.	0.99 (0.82-1.19)	1.10 (0.88-1.38)
Model 1 [e]	**HR Per 1 SD Increase (95% CI)**		0.94 (0.89-1.00)			1.06 (0.94-1.18)		
	P-trend		0.050			0.354		
	HR (95% CI)	0.87 (0.69-1.10)	Ref.	0.92 (0.81-1.04)	0.91 (0.78-1.07)	Ref.	0.93 (0.76-1.14)	1.11 (0.87-1.41)
Model 2 [e, f]	**HR Per 1 SD Increase (95% CI)**		0.94 (0.89-1.01)			1.06 (0.93-1.20)		
	P-trend		0.072			0.380		

[e] Model 1: Adjusted for age (years, continuous), sex (male or female, if applicable), smoking (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), and total energy intake (kcal/day, continuous).

[e, f] Model 2: Additionally, ethnicity (Caucasian or non-Caucasian, if applicable) , alcohol intake (ml/day, continuous), fruit intake (g/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), and vegetable intake (g/day, continuous).

Abbreviation: HR, hazard ratio; CI, confidence interval; SD, standard deviation; MIBC, muscle invasive bladder cancer; NMIBC, non-muscle invasive bladder cancer; g, gram; kcal, kilocalorie; ml, millilitre.

The intervals of tertiles were categorized as open lower values and ended upper values.

Reference group was lowest intake (tertile 1).

P-trend <0.05 was considered statistically significant.

Supplementary Table 6-5 *Risk of Bladder Cancer According to Intake of Total Whole Grain, Total Refined Grain and Total Dietary Fibre in Different Adjustment Models*

Model Adjustments		Model 1) [g]	Model 2) [h]	Model 3) [i]
	Case/Total	1,733/217,556	1,211/146,385	1,211/146,385
Total Whole Grain (g/day)	Low (0-3)	Ref.	Ref.	Ref.
	Medium (3-8)	0.98 (0.86-1.11)	0.99 (0.87-1.12)	0.97 (0.85-1.10)
	High (>8)	0.83 (0.73-0.94)	0.84 (0.74-0.95)	0.81 (0.72-0.92)
	Case/Total	3,214/574,726	2,228/418,548	2,228/418,548
Total Refined Grain (g/day)	Low (0-100)	Ref.	Ref.	Ref.
	Medium (100-181)	0.93 (0.84-1.02)	0.93 (0.85-1.03)	0.93 (0.85-1.03)
	High (>181)	0.94 (0.84-1.05)	0.95 (0.84-1.06)	0.95 (0.85-1.06)
	Case/Total	3,214/574,726	2,228/418,548	2,228/418,548
Total Dietary Fibre (g/day)	Low (0-15)	Ref.	Ref.	Ref.
	Medium (15-23)	0.92 (0.83-1.01)	0.93 (0.84-1.03)	0.93 (0.84-1.03)
	High (>23)	0.86 (0.77-0.96)	0.87 (0.78-0.98)	0.88 (0.79-0.97)

Adjustments: age (years, continuous), sex (male or female), smoking (if applicable, smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), total energy intake (kcal/day, continuous), smoking status (never, former, current), and smoking pack-years (continuous).

[g] Model 1): adjusted for age, sex, energy intake and smoking status.

[h] Model 6): adjusted for age, sex, energy intake and smoking pack-years.

[i] Model 6): adjusted for age, sex, energy intake, smoking status and smoking pack-years.

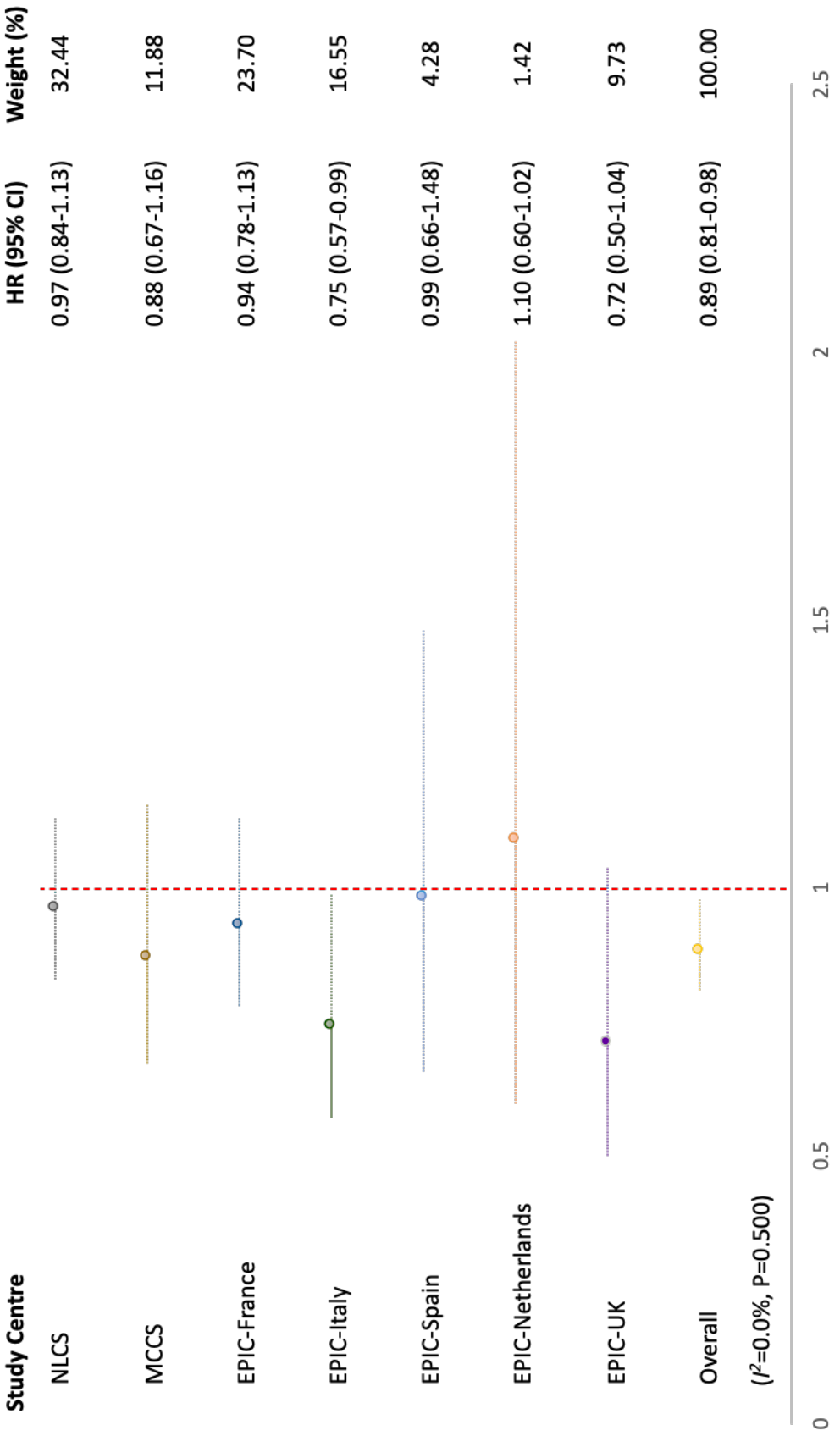
Abbreviation: HR, hazard ratio; CI, confidence interval; g, gram; kcal, kilocalorie; ml, millilitre.

The intervals of tertiles were categorized as open lower values and ended upper values.

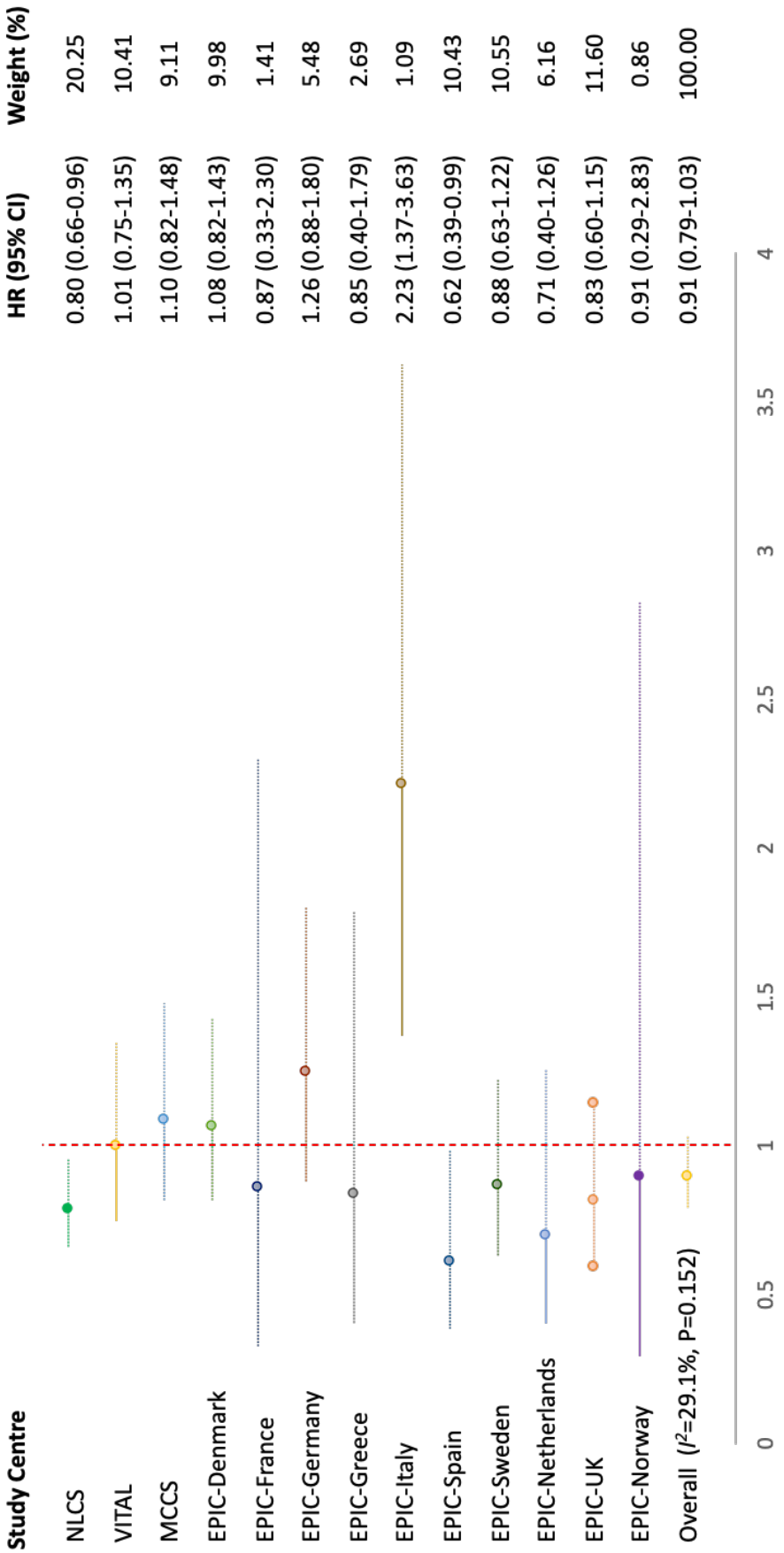
Reference group was lowest intake (tertile 1).

P-trend <0.05 was considered statistically significant.

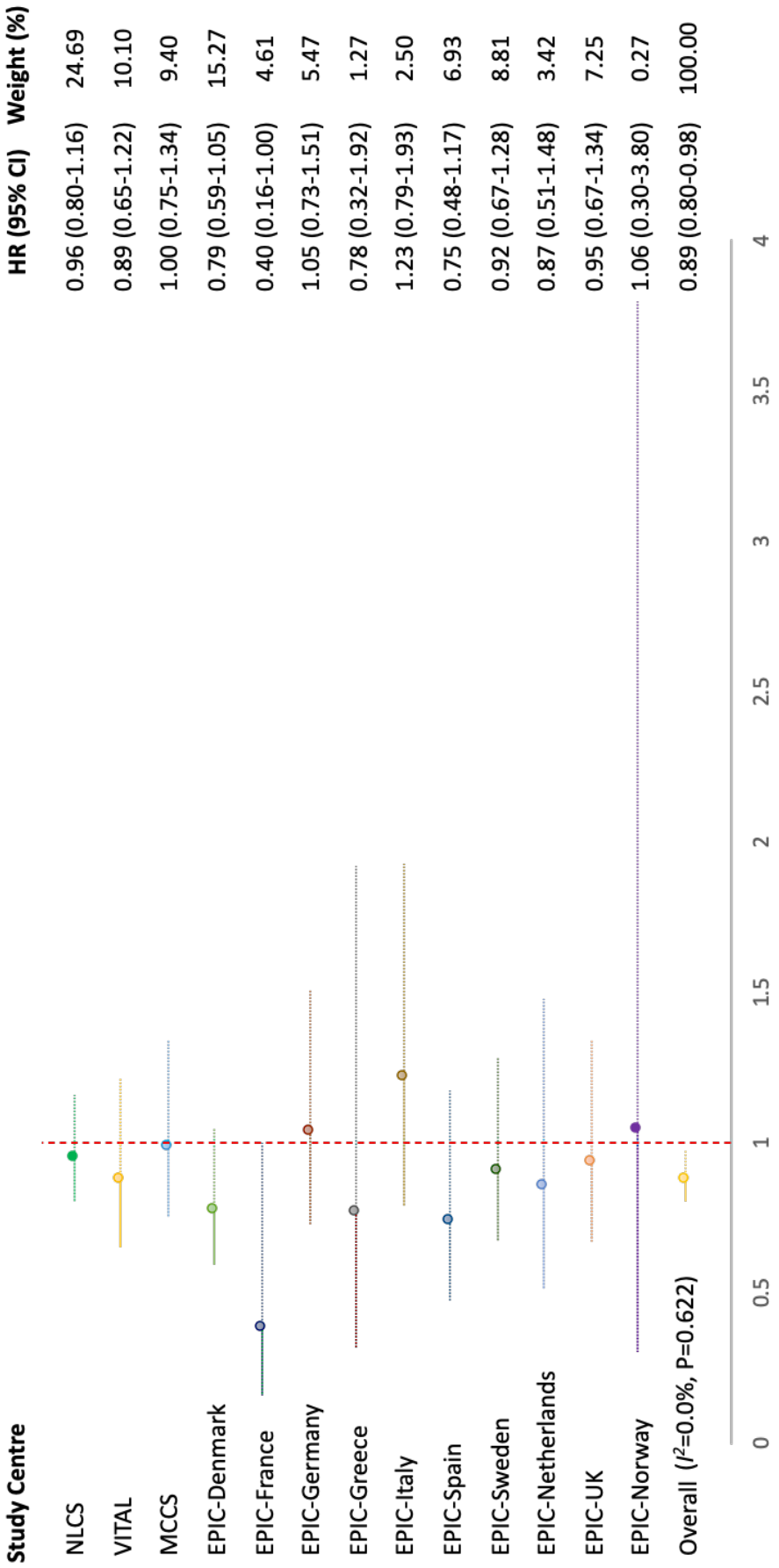
a) Total Whole Grain



b) Total Refined Grain



c) Total Dietary Fibre



Supplementary Figure 6-1 *Forest Plot of Meta-Analyses with HRs and 95% CIs for High vs. Low Intake of a) Total Whole Grain, b) Total Refined Grain and c) Total Dietary Fibre with Bladder Cancer Risk*

Diamond dots denote the hazard ratios (HRs); Horizontal lines represent the 95% confidence intervals (CIs); Weights (grey squares) are from random effects analyses.
Adjusted for age (years, continuous), sex (male or female), smoking (smoking was defined as: 0 (never smokers); 1 [current light smokers (i.e. smoking less than 20 pack-years)]; 2 [current heavy smokers (i.e. smoking more than 20 pack-years)]; 3 [current smokers (no information on pack-years)]; 4 [former light smokers (i.e. smokers who ceased smoking over 1 year prior and smoked less than 20 pack-years)]; 5 [former heavy smokers (i.e. smokers who ceased smoking over 1 year prior and smoked more than 20 pack-years)]; 6 [former smokers (smokers who ceased smoking over 1 year prior and no information on pack-years)]), total energy intake (kcal/day, continuous), ethnicity (Caucasian or non-Caucasian, if applicable) , alcohol intake (ml/day, continuous), fruit intake (g/day, continuous), fat intake (g/day, continuous), meat intake (g/day, continuous), sugar intake (g/day, continuous), and vegetable intake (g/day, continuous).
Abbreviation: HR, hazard ratio; CI, confidence interval; g, gram; kcal, kilocalorie; ml, millilitre.
Reference group was lowest intake (tertile 1).
P-value <0.05 was considered statistically heterogeneous.

Chapter 7

A Machine Learning Approach to Investigate Food Groups related to Incidence of Bladder Cancer based on 18 Case-control Studies and 1 Nested Cohort Study

Evan Y.W. Yu, Anke Wesselius, Christoph Sinhart, Alicja Wolk, Mariana Carla Stern, Xuejuan Jiang, Li Tang, James Marshall, Eliane Kellen, Piet van den Brandt, Chih-Ming Lu, Hermann Pohlabeln, Gunnar Steineck, Mohamed Farouk Allam, Margaret R. Karagas, Carlo La Vecchia, Stefano Porru, Angela Carta, Klaus Golka, Kenneth C. Johnson, Simone Benhamou, Zuo-Feng Zhang, Jerry Polesel, Cristina Bosetti, Jack A. Taylor, Elisabete Weiderpass, Eric J. Grant, Emily White, Maurice P.A. Zeegers

Abstract

Background At present, the analysis of diet and bladder cancer is mostly based on the intake of individual foods. The examination of food combinations provides a scope to deal with the complexity and unpredictability of the diet and aims to overcome the limitations of the study of nutrients and foods in isolation.

Objective This article aims to demonstrate the usability of supervised machine learning methods to extract the food groups related to bladder cancer.

Design A pooled international consortium [*BLadder cancer Epidemiology and Nutritional Determinants* (BLEDN) study] with 18 case-control studies and 1 nested case-cohort study was performed.

Participants/setting Participants included 8,320 bladder cancer cases and 31,551 non-cases within the age range of 18-100 years originated from 12 different countries across the world.

Main outcome measure Five key food groups with sex, smoking and age showed to be associated with bladder cancer risk.

Statistical analyses A classification technique called C5.0 with built-in "ruleset" function was used in order to derive key food groups associated with bladder cancer risk.

Results: Primarily, five key food groups were extracted; in order of importance: beverages (non-milk); grains and grain products; vegetables and vegetable products; fats, oils and their products; meats and meat products were related bladder cancer risk.

Conclusion: Since these food groups are corresponded with previously proposed bladder cancer related dietary factors, machine learning seems to be a promising technique in the field of nutritional epidemiology and deserves further examination.

Key Words: Bladder Cancer; Machine Learning; Food Groups; Epidemiological Studies

Introduction

Bladder cancer is the most common malignancy of urinary tract and the seventh cause of mortality for cancer (2.8% of all cancer deaths), with nearly 430,000 new cases and 165,000 deaths per year worldwide [1, 2]. According to *Al-Zalabani et al*, up to 80% of bladder cancer can be attributed to lifestyles, including occupation, smoking, exercise and diet [3]. Particularly, it is biologically plausible for dietary factors to influence bladder cancer risk considering that beneficial as well as harmful components of diet are excreted through the urinary tract and in direct contact with the epithelium of the bladder [4]. However, as stated in the report by WCRF/AIRC [5], there is still "limited" evidence for the role of diet on the bladder cancer risk.

Analysis of overall dietary patterns related to bladder cancer has been gained a lot of attention during past years [6, 7]. Instead of looking at individual foods or nutrients, analysis of dietary patterns examines the effects of the overall diet, considering the inter-correlations in the consumption of various foods and nutrients. Conceptually, dietary patterns represent a broader picture of food and nutrient consumption, and analysis of dietary patterns may help in better understanding and preventing the development of common cancers.

Several conventional analysis techniques are available for extracting dietary patterns including factor and cluster analyses; a) investigator-driven methods, such as dietary indices and dietary scores; b) and data-driven methods, such as *principle component analysis* (PCA). Although these techniques are widely used and might reveal some important information on the relation between dietary patterns and common cancers, they all draw subjective conclusions since they are based on series of priori assumptions, which may differ among researchers. A relatively new approach in the field of nutritional epidemiology is "machine learning". Machine learning is a process that uses a variety of data analysis tools to extract hidden predictive information from large data. This technique is considered to be a powerful technology with great potential to help people focus on the most important information of their data [8]. A previous study in the field of nutritional epidemiology already showed that machine learning allowed to define unexpected dietary patterns that might not be recognized using conventional statistical methods [9]. Therefore, in the present study we used this

technique to examine the combinational foods at individual level to extract some food groups related to the bladder cancer risk.

Methods

Study Population

The dataset used in the present study is part of the "BLadder cancer Epidemiology and Nutritional Determinant (BLEND)" study, which aims at assessing the association between diet and the bladder cancer risk. Details on the methodology of the BLEND consortium have been described elsewhere [10]. The present study included data of 18 case-control studies [11-28] and 1 nested case-cohort study [29] providing information on diet and bladder cancer, from 12 different countries across the world, including data on 8,320 bladder cancer cases and 23,231 non-cases within the age range of 18-100 years. Each study ascertained incident bladder cancer, defined to include all urinary bladder neoplasms according to the International Classification of Diseases for Oncology (ICD-O-3 code C67) using population-based cancer registries, health insurance records, or medical records. Each participating study has been approved by the local ethic committee. Informed consent was obtained from all individual participants included in each study. Most of the bladder cancer cases were diagnosed and histologically confirmed in 1990s.

Data Collection

All included studies made use of a validated self-administrated *food frequency questionnaire* (FFQ) or an FFQ administered by a trained interviewer. Homogenization of the dietary data was done by making use of the Eurocode 2 Core classification codebook [30]. This codebook consists of main food groups and their first and second-level subgroups [31]. In order to reduce the variance of individual food items across the world (**Supplementary Table 7-1**), foods were attributed into 11 main groups: milk and dairy products (A); eggs and egg products (B); meats and meat products (C); fishes and fish products (D); fats, oils and their products (E); grains and grain products (F); pulses, seeds, kernels, nuts and their products (G); vegetables and vegetable products (H); fruits and fruit products (I); sugars and sugar products (J); beverages (non-milk, K). All food groups were measured as servings of food intake per week and divided into quartile, with Q1-Q4 respectively corresponding to lowest- and highest intake). In addition to information on diet, the BLEND dataset also included data on study

characteristics (design, method of dietary assessment, and geographical region), participant demographics [(age (continuous), sex (male, female)] and smoking status (never/current/former).

Baseline analysis

Continuous variables were described as mean and *standard deviation* (SD), and categorical variables as absolute and relative frequencies. Missing values were tested for *missing at random* (MAR) or *missing completely at random* (MCAR) [32, 33]. To test for MAR, logistic regression was performed with a missing data indicator created for each variable. No significant relationship between the missingness indicators and the outcome of interest suggests MAR. The assumption that missing data are MCAR were assessed using the Little's MCAR chi-squared test [34, 35].

Machine Learning Method

All the 11 main food groups and the non-diet variables (*i.e.* age, sex and smoking status) were selected and entered into machine learning procedures.

A classification technique called C5.0 [36], which is a variant of the C4.5 algorithm developed by Ross Quinlan, was used since it can represent solutions as decision trees and as rulesets [37]. It builds a decision tree based on the training/validation sets using the concept of information entropy. The decision tree is built by splitting the data in two parts at the value of one variable that yields the highest normalized information gain. That is, it splits on the value of the chosen variable that separates positive and negative observations (*i.e.* bladder cancer status; case and non-case), most efficiently. The pruning severity of the model was set at the default level of 75. This level yielded the lowest complexity (*i.e.* which refers to the minimum number of records in each tree branch to allow a split) with sufficient accuracy. Standard tenfold cross-validation was used in which the entire eligible BLEND dataset was divided into ten approximately equally sized parts. Nine parts were used in turn as training sets and the remaining tenth part was used as the validation set. The validation set (10%) was chosen within the entire dataset according to the distribution of bladder cancer status. The participants with missing values were taken into account by using the ratio of the participants with missing values multiplied by the information entropy of the subset of participants

without missing values for each variable [38]. The classification C5.0 algorithm was run for the included diet and non-diet variables within the BLEND dataset; meanwhile, variable importance (*i.e.* attribute usage) for C5.0 model was calculated by determining the percentage of training set samples that fall into all the terminal nodes after the split, which defines the variable importance value of each diet and non-diet variables in relation to bladder cancer [39-42]. These importance values range from 0% to 100%, where 0% indicates "unimportant" and 100% indicates "extremely important". Both continuous and categorical variables were included in the models. Node splits in continuous variables can occur at any value and were not predetermined.

Rules were then generated by using the "ruleset" function in C5.0, which transformed the decision tree into specific context associated with bladder cancer. A bladder cancer status (either case or non-case) was predicted by each rule, and a value between 0% and 100% indicates the confidence of the risk in relation to bladder cancer outcome. The overall performance of the C5.0 classifier was evaluated by classification accuracy, *true positive rate* (TPR), *false positive rate* (FPR) and *receiver operating characteristic* (ROC) with the *area under the ROC curve* (AUC). This is the number of correct classifications of the instances from the validation set divided by the total number of these instances, expressed as a percentage. The greater the classification accuracy, the better is the classifier. A sensitivity analysis was performed by categorizing age in to six groups (years): ≤55, 55-60, 60-65, 65-70, 70-75, >75, based on same machine mining procedure.

All data analyses were performed with R software (version 3.5.1). (using packages "C5.0" and "caret" developed by Max Kuhn; "rpart" developed by Beth Atkison; "ROCR" developed by Tobias Sing and Oliver Sander).

Results

Baseline analyses of the included data

The characteristics of the BLEND participants are presented in **Table 7-1**. In total 31,551 participants are included in the analyses, of which 8,320 (26.4%) were bladder cancer cases. The mean age of non-cases (59 years old) was lower than cases (62 years old), and most of

the participants were Caucasian (92.3%). Around 66.7% of participants were smokers, with 33.62% of those being current smokers and 33.06% being former smokers.

Table 7-1 *The Baseline Characteristics and Food Group Information of Included Studies* [a]

Variables	Cases (*N*=8,320)	Non-cases (*N*=23,231)	Missing Percentage
Sex (%)			0.00%
Male	6,601 (33.95)	12,841 (66.05)	
Female	1,719 (14.20)	10,390 (85.80)	
Smoking (%)			0.00%
Never	1,588 (15.11)	8,925 (84.89)	
Current	3,285 (30.97)	7,321 (69.03)	
Former	3,447 (33.04)	6,985 (66.96)	
Age (± SD)	61.80 (± 10.61)	58.52 (± 12.54)	0.00%
≤55 (%)	1,880 (20.46)	7,310 (79.54)	
55-60 (%)	1,511 (26.67)	4,514 (73.33)	
60-65 (%)	1,708 (28.22)	4,345 (71.78)	
65-70 (%)	1,531 (27.97)	3,943 (72.03)	
70-75 (%)	1,068 (31.23)	2,352 (68.77)	
>75 (%)	622 (35.56)	1,127 (64.44)	
Main Food Groups [mean servings/week (± SD)]			
Milk and Milk Products (± SD)	13.61 (± 18.39)	14.58 (± 23.47)	4.47%
Q1 (%): 0-5 servings/week	1,887 (22.14)	6,632 (79.86)	
Q2 (%): 5-9 servings/week	1,322 (19.93)	5,310 (80.07)	
Q3 (%): 9-18 servings/week	1,626 (21.81)	5,828 (78.19)	
Q4 (%): >18 servings/week	1,500 (19.92)	6,031 (80.08)	
Eggs and Egg Products (± SD)	2.65 (± 2.89)	2.54 (± 2.63)	11.78%
Q1 (%): 0-1 servings/week	2,117 (22.87)	7,141 (77.13)	
Q2 (%): 1-2 servings/week	1,003 (18.89)	4,306 (81.11)	
Q3 (%): 2-3 servings/week	1,246 (17.55)	5,852 (82.45)	
Q4 (%): >3 servings/week	1,275 (26.05)	4,894 (73.95)	

Table 7-1_continued *The Baseline Characteristics and Food Group Information of Included Studies* [a]

Variables	Cases (*N*=8,320)	Non-cases (*N*=23,231)	Missing Percentage
Meat and Meat Products (± SD)	7.75 (± 5.54)	7.35 (± 4.47)	7.62%
Q1 (%): 0-5 servings/week	1,931 (24.41)	5,981 (75.59)	
Q2 (%): 5-8 servings/week	1,810 (22.73)	6,154 (77.27)	
Q3 (%): 8-11 servings/week	1,387 (21.32)	6,505 (78.68)	
Q4 (%): >11 servings/week	1,298 (16.53)	6,555 (83.47)	
Fish and Fish Products (± SD)	1.94 (± 2.08)	1.39 (± 1.73)	5.72%
Q1 (%): 0-0.5 servings/week	918 (24.41)	7,415 (75.59)	
Q2 (%): 0.5-1 servings/week	1,163 (12.02)	8,515 (87.98)	
Q3 (%): 1-2 servings/week	1,387 (17.45)	4,287 (82.55)	
Q4 (%): >2 servings/week	1,298 (19.10)	5,293 (80.90)	
Fats and Oils (± SD)	8.61 (± 7.98)	9.99 (± 8.77)	21.44%
Q1 (%): 0-4 servings/week	1,291 (20.53)	5,641 (79.47)	
Q2 (%): 4-7 servings/week	1,561 (23.79)	5,760 (76.21)	
Q3 (%): 7-10 servings/week	780 (13.49)	5,386 (86.51)	
Q4 (%): >10 servings/week	1,152 (18.73)	5,654 (81.27)	
Grains and Grain Products (± SD)	16.17 (± 17.33)	15.40 (± 15.67)	5.36%
Q1 (%): 0-7 servings/week	2,061 (25.80)	5,928 (74.20)	
Q2 (%): 7-13 servings/week	1,503 (21.63)	5,446 (78.37)	
Q3 (%): 13-21 servings/week	1,494 (20.01)	5,973 (79.99)	
Q4 (%): >21 servings/week	1,570 (21.06)	5,886 (78.94)	
Pulses, seeds, kernels and nuts (± SD)	2.68 (± 3.88)	2.98 (± 4.44)	31.41%
Q1 (%): 0-0.75 servings/week	766 (13.89)	4,747 (86.11)	
Q2 (%): 0.75-1.5 servings/week	671 (12.06)	4,895 (87.94)	
Q3 (%): 1.5-3 servings/week	631 (12.24)	4,523 (87.76)	
Q4 (%): >3 servings/week	632 (11.69)	4,776 (88.31)	
Vegetables and Vegetable Products (± SD)	29.48 (± 47.94)	26.53 (± 34.52)	4.47%
Q1 (%): 0-12 servings/week	1,992 (26.29)	5,585 (73.71)	
Q2 (%): 12-17 servings/week	1,629 (21.74)	5,864 (78.26)	

Table 7-1_continued *The Baseline Characteristics and Food Group Information of Included Studies* [a]

Variables	Cases (*N*=8,320)	Non-cases (*N*=23,231)	Missing Percentage
Q3 (%): 17-29 servings/week	1,422 (18.86)	6,118 (81.14)	
Q4 (%): >29 servings/week	1,590 (21.12)	5,940 (78.88)	
Fruits and Fruit Products (± SD)	9.18 (± 9.97)	10.74 (± 12.59)	8.16%
Q1 (%): 0-3 servings/week	2,056 (26.46)	5,715 (73.54)	
Q2 (%): 3-6 servings/week	1,100 (14.79)	6,338 (85.21)	
Q3 (%): 6-14 servings/week	2,019 (28.55)	5,052 (71.45)	
Q4 (%): >14 servings/week	1,281 (18.69)	5,574 (81.31)	
Sugar and Sugar Products (± SD)	10.99 (± 14.67)	7.07 (± 10.65)	30.52%
Q1 (%): 0-1 servings/week	667 (10.40)	5,745 (89.60)	
Q2 (%): 1-4 servings/week	438 (9.37)	4,236 (90.63)	
Q3 (%): 4-10 servings/week	641 (11.95)	4,725 (88.05)	
Q4 (%): >10servings/week	896 (16.38)	4,573 (83.62)	
Beverages (non-milk) (± SD)	56.84 (± 17.63)	45.53 (± 13.47)	4.20%
Q1 (%): 0-28 servings/week	2,399 (23.90)	7,083 (76.10)	
Q2 (%): 28-42 servings/week	1,491 (23.86)	4,579 (76.14)	
Q3 (%): 42-62 servings/week	1,696 (24.15)	5,328 (75.85)	
Q4 (%): >62 servings/week	2,570 (34.40)	4,901 (65.60)	

[a] *Age was coded as the original continuous values and 6 categorical values, food intakes were coded as quartile-order categorical values, and the other variables were coded as categorical dummy values.*
Q1-Q4: lowest intake to highest intake (servings/week).
Abbreviation: SD, Standard Deviation.

Significant results of logistic regression for food-group variables indicated that missing dietary data was not MAR (all $P\text{-}_{MAR}$ <0.05). Little's test also provided evidence against the assumption that missing data were MCAR (all $P\text{-}_{MCAR}$ <0.001). Rejection of both MAR and MCAR indicates the missing values are missing not at random (MNAR). Therefore, the observations with missing data could not be deleted, and the missing values were marked as blank and not replaced by any value.

Extraction of food groups in relation to bladder cancer via the machine learning procedure

Figure 7-1 *presents an example of a decision tree with three different variables. The variables are ranked according to how they were used to split the participants from decision nodes to end nodes. A position of 1 (A) corresponds to the variable that in all trees is the first variable used to split; a position of 2 (B) corresponds to the variable that on average is the second variable used to spit, and so on till finally the all participants were split into BC cases and non-cases.* "Sex" is on the first rank split of the tree, which indicates dietary patterns are differentiated in males and females related to bladder cancer. Both non-diet variables (age, sex and smoking status) and five food groups (C, E, F, H, K) were identified as having an influence on development of bladder cancer. The observed importance values of these variables are (**Figure 7-2**): sex (100%); smoking status (74.60%); age (62.80%); beverages (37.98%); grains and grain products (55.81%); vegetables and vegetable products (24.30%); fats, oils and their products (2.95%); meats and meat products (2.17%). Other input variables showed to have an importance value of 0% and were, therefore, considered non-relevant for BC development. The overall classification accuracy is 75.10%, with TPR 0.86 and FPR 0.31 (the ROC curves, with AUCs from 0.690-0.701, for each cross-validation run were performed in **Supplementary Figure 7-1**).

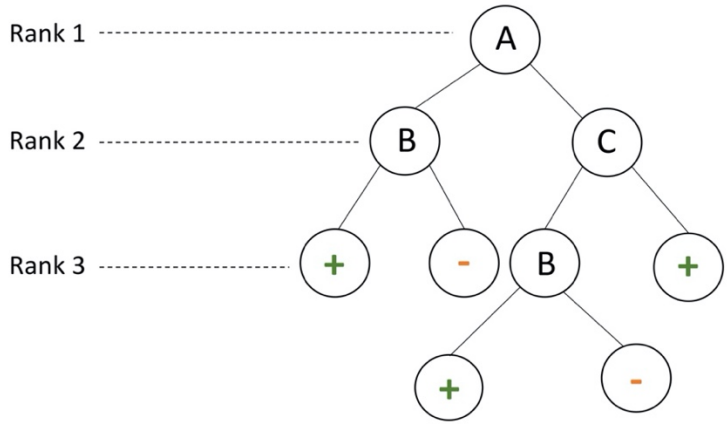


Figure 7-1 *An example of a decision tree.*
There are three individual variables, A, B, and C, on which the tree splits. Variable A has an average ranking of 1 because it is the root node and appears only once. Variable B has an average ranking of 2.5, since it appears twice. Once on second and once on the third rank. Variable C has an average ranking of 2, since it is present only once and the tree splits on it after it split on A.

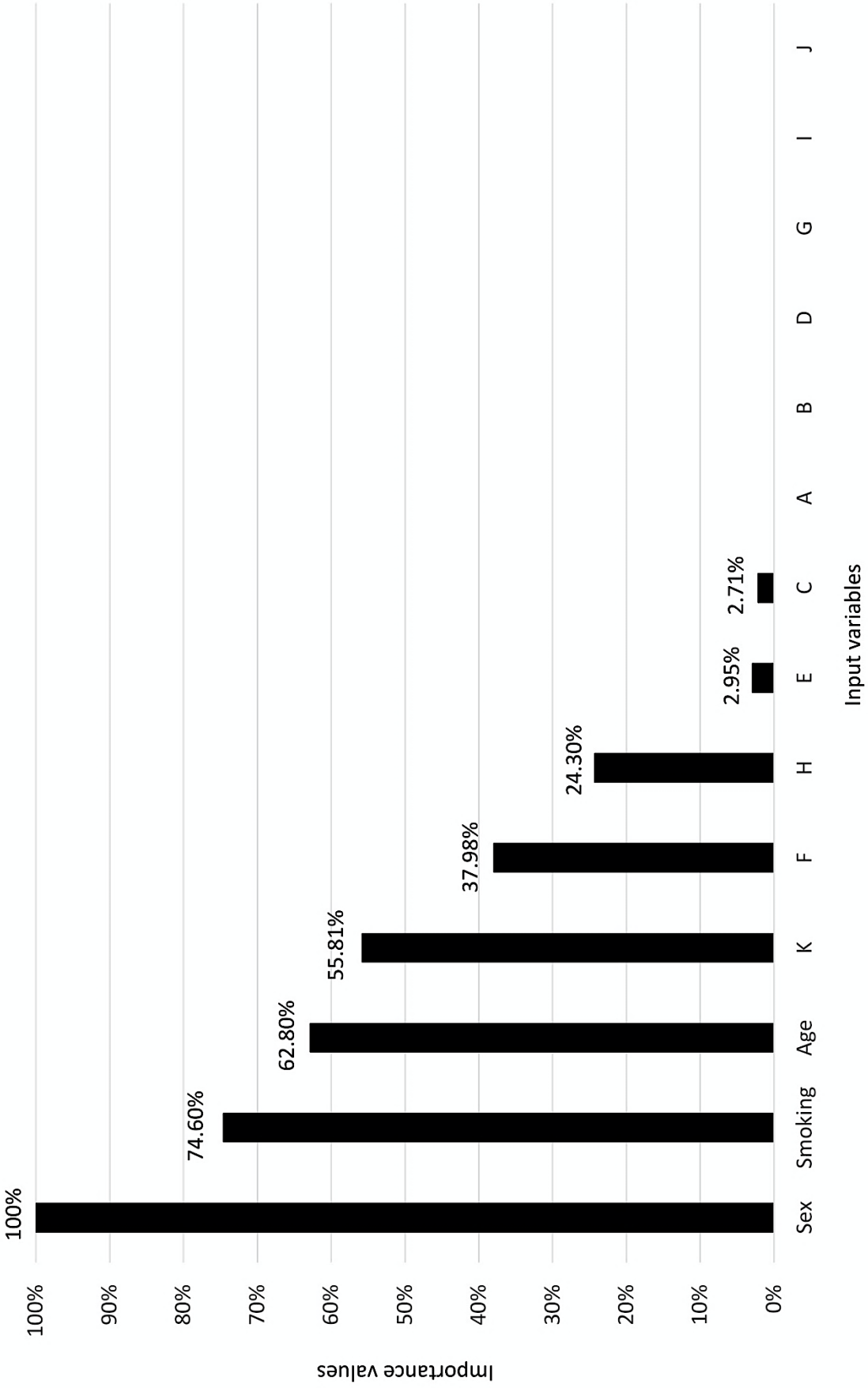

100%
74.60%
62.80%
55.81%
37.98%
24.30%
2.95%
2.71%
Sex
Smoking
Age
K
F
H
E
C
A
B
D
G
I
J
Input variables
Importance values
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%

Figure 7-2 *Importance Values of Input Variables after C5.0*
A: milk and dairy products; B: eggs and egg products; C: meats and meat products; D: fishes and fish products; E: fats, oils and their products; F: grains and grain products; G: pulses, seeds, kernels, nuts and their products; H: vegetables and vegetable products; I: fruits and fruit products; J: sugar and sugar products; K: beverages (non-milk).
The importance values range from 0% to 100%, where 0% indicates "unimportant" and 100% indicates "extremely important".

Table 7-2 presents the extracted eight rules resulting into BC outcome after application of the "ruleset" classifier of C5.0, with a classification accuracy of 74.90%. The results from "ruleset" show that the variables identified by the "decision tree" approach are also identified by using the "ruleset" approach. Here we see that current/former males smokers tended to be BC cases and never male smokers tended to be non-BC cases. However, to be able to split the participants into case or non-case is depending on their dietary habits. Females show relatively simple rules, in which "only grain and grain products" and "beverages (non-milk)" were identified to be related to BC.

Table 7-2 *Classification Rules Derived from C5.0 "Ruleset"* [b]

Sex	Rules	Age	Smoking status	C	E	F	H	K	Case (%)	Non-case (%)
Male	1		Current		Q1	Q3-Q4	Q3-Q4	Q1-Q2	24%	76%
	2	40-63	Former			Q1		Q4	67%	33%
	3			Q3-Q4		Q1	Q1-Q2		87%	13%
	4		Never						23%	77%
	5		Current/ Former						63%	37%
Female	6					Q2-Q4			13%	87%
	7		Former			Q1		Q1-Q2	80%	20%
	8	>63	Former			Q1		Q3-Q4	84%	16%

[b] *Age: years old; C-K: servings/week.*
C: meats and meat products; E: fats, oils and their products; F: grains and grain products; H: vegetables and vegetable products; K: beverage (non-milk).
Q1-Q4: lowest intake to highest intake (servings/week).

A sensitivity analysis by transforming age into categorial variable was performed based on C5.0 algorithm, the results show similar with identification of same food groups related to BC (**Supplementary Figure 7-2**).

Discussion

To our knowledge, this is among the first studies to apply the machine learning approach to extract food groups associated with bladder cancer risk based on the complexity of the combinational food intake. By applying C5.0 algorithm, the decision tree and rules derived from this approach showed that sex, smoking status, age and five food groups [C: meats and meat products, E: fats, oils and their products, F: grains and grain products, H: vegetables and vegetable products, K: beverages (non-milk)] are in relation with bladder cancer risk in both males and females. Apart from the well-established factors (*e.g.* age, sex and smoking) for bladder cancer identified in the machine learning procedures, the association of diet, especially specific dietary pattern, with bladder cancer risk deserves to be explored due to the limited evidence on this topic, and because it reflects a person's dietary exposure in aggregate rather than in isolation.

Although the use of machine learning is relatively new for unravelling diet in relation to the cancer risk, previous studies already examined dietary intake with bladder cancer risk using other techniques. In 2008, *De Stefani et al.* [43] found that the dietary patterns labelled as "sweet beverages" (high loadings of coffee, tea, and added sugar) and "Western" (high loadings of red meat, fried eggs, potatoes, and red wine) were directly associated with the risk of bladder cancer based on factor analysis. In addition, the negative influence of the Western diet was also observed for bladder cancer recurrence: bladder cancer patients in the highest tertile of adherence to a Western dietary pattern had a 48% higher risk of recurrence of bladder cancer compared to patients in the lowest tertile [6]. The Western diet is especially low in fresh fruits and vegetables, but generally high in saturated fats and red and processed

meats. Results from the present study are in line with these results, with respect to high of fat being associated with an increased risk for the development of bladder cancer and high intake of vegetables and vegetable products being associated with a reduced risk.

Previous studies on single food item or food groups in relation to bladder cancer risk, also reported that high intake of vegetables was associated with reduced risk of bladder cancer [44-47]. These studies suggest that the preventive effect could possibly be due to the antioxidant action of vegetables [48, 49] and that each serving of vegetable may result in a 10% risk decline. Although very powerful, results from the present study only identify "vegetables and vegetable products" as a possible main food group related to bladder cancer risk. It remains unclear which specific subgroup is responsible (*e.g.* starchy/non-starchy, processed/fresh, citrus/cruciferous). Detailed analyses of BLEND data may help to elucidate this uncertainty.

Limited evidence is available on the influence of "grains and grain products" on bladder cancer risk. However, our findings are in line with results from a previously conducted case-control [50], suggesting that a high intake of whole grains may reduce the risk of bladder cancer. In contrast, a more recent study found that bladder cancer risk was negatively influenced by a high intake of refined carbohydrate foods [51]. Thus, future detailed analyses, especially those focusing on whole grains and refined grain products, may be useful. Of note, our results on grain products might have been influenced by the fact that the 'grain and grain products' group of the present study included sweet 'Fine bakery wares' such as 'Sweet biscuits and cookies' which are high in sugar and thereby promote obesity, known to be a risk factor for BC [52].

Only few studies discussed the associations between fat, oil and their products and bladder cancer risk and were summarized in a systematic review. This review showed that the total fat intake was positively related to bladder cancer risk when combing results from three case-control studies. However, no such association was observed in cohort studies [53]. The present study confirms findings from the case-control studies, in that a positive association was found.

A meta-analysis reported that overall meat intake was not related to the risk of bladder cancer; however, high red and processed meat intake was reported as a significant risk factor for bladder cancer risk, 17% and 10% risk respectively [54]. This increase is probably caused by the N-Nitroso compounds, which have been proposed as possible bladder carcinogens, found in red and processed meats [55]. In the present study, a high intake of 'meats and meat products (C)' was associated with an increased risk of developing BC. Again, future studies investigating specific types of meat could identify the types of meat or meat products that might have beneficial effects.

As an excretory organ, fluid intake might play an important role in the development of bladder cancer. A well-established risk factor is arsenic [56], through which people are most likely exposed by drinking water. The influence of other fluid sources on bladder cancer risk, however, are lacking evidence or are inconstant. Here we observed that high beverage intake is positively associated with bladder cancer risk. Again, it should be noted that only total "beverage" intake was assessed, including both beverages with a potential protective effect on BC risk (e.g. green tea [57]) and beverages with a potential harmful effect on BC risk (e.g. alcoholic [58] and sweet non-alcoholic beverages [43]). It, therefore, remains unclear which caused the observed increased BC risk.

Since nutrition and cancer epidemiology is a complex field, the use of advanced analytic tools, such as machine learning, is becoming increasingly important for unrevealing diet and health associations. Machine learning has demonstrated its potential to complement conventional statistical regressions, particularly for nonlinear phenomena such as our dietary habits [59], and without requiring a priori assumptions on the relationship between diet and health outcomes [60]. In addition, machine learning splits data files into training and validation sets, especially using cross-validation method gives relatively accurate predictive estimates. Furthermore, over fitting problem of both decision tree and rules could be minimized by using reduced error pruning technique in C5.0 [36] which is often problematic in conventional statistical techniques with a large number of variables and observations, such as the BLEND dataset. The strength of the present study is the high classification accuracy, which indicates the machine learning methodology could adequately handle missing data and complex-

investigating measurements. Therefore, the revealed food groups in the present study could be considered foods or pattern in relation to bladder cancer development.

A limitation of our study, however, is that the use of machine learning in nutritional cancer epidemiology might only be useful in identifying key food items and can therefore only be seen as a hypothesis generator, which needs further detailed investigation in order to establish causation. Furthermore, we should acknowledge it is a complicated technique, which requires special knowledge and expertise and, thus, translating the results from machine learning into simple health message is difficult challenge. In addition, the trees and rules retrieved in the present study only include main food-groups, thereby, conflicting effects on bladder cancer risk of food subgroups or specific items was inevitable. Another limitation might have occurred by the designs of the data collection, which may have introduced recall- and/or selection bias, especially in case-control studies. In addition, for most included studies, the exposure variable was assessed by FFQs. Therefore, measurement error and misclassification of study participants in terms of the exposure and outcome are unavoidable: a) the inability of an FFQ to capture many details of dietary intake, such as all kinds and exact amounts of foods consumed, b) the difficulty in quantification of the intake and c) the high dependency on memory, which in turn may have influenced the robustness of dietary patterns extracted via the machine learning procedure [61]. Lastly, due to the nature of machine learning such as C5.0, there are concerns regarding multiple testing and spurious associations, which might cause some of the observed consequences due to chance alone.

Conclusion

In summary, the machine learning technique provided an effective approach to identify some food groups related to bladder cancer risk in the large epidemiological BLEND study. The main findings from this study support the machine learning approach to be a valuable additional methodology in nutrition- and cancer epidemiology, which deserve further examines.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer. 2015;136(5):E359-86. doi:10.1002/ijc.29210.

2. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. *CA Cancer J Clin* 2017;67(1):7-30. doi:10.3322/caac.21387.

3. Al-Zalabani AH, Stewart KF, Wesselius A, Schols AM, Zeegers MP. Modifiable risk factors for the prevention of bladder cancer: a systematic review of meta-analyses. European journal of epidemiology. 2016;31(9):811-51. doi:10.1007/s10654-016-0138-6.

4. Piyathilake C. Dietary factors associated with bladder cancer. Investigative and clinical urology. 2016;57 Suppl 1:S14-25. doi:10.4111/icu.2016.57.S1.S14.

5. Wiseman M. World Cancer Research Fund/American Institute for Cancer Research. Diet, Nutrition, Physical Activity and Cancer: a Global Perspective. Continuous Update Project Expert Report 2018.

6. Westhoff E, Wu X, Kiemeney LA, Lerner SP, Ye Y, Huang M et al. Dietary patterns and risk of recurrence and progression in non-muscle-invasive bladder cancer. *Int J Cancer* 2018;142(9):1797-804. doi:10.1002/ijc.31214.

7. Witlox WJA, van Osch FHM, Brinkman M, Jochems S, Goossens ME, Weiderpass E et al. An inverse association between the Mediterranean diet and bladder cancer risk: a pooled analysis of 13 cohort studies. European journal of nutrition. *EUR J NUTR* 2019:1-10. doi:10.1007/s00394-019-01907-8.

8. Han J, Kamber M. Data mining concept and technology. *Publishing House of Mechanism Industry* Amsterdam Elsevier Inc. 2001;3rd:70-2.

9. Hearty AP, Gibney MJ. Analysis of meal patterns with the use of supervised data mining techniques--artificial neural networks and decision trees. *Am J Clin Nutr* 2008;88(6):1632-42. doi:10.3945/ajcn.2008.26619.

10. Goossens ME, Isa F, Brinkman M, Mak D, Reulen R, Wesselius A et al. International pooled study on diet and bladder cancer: the bladder cancer, epidemiology and nutritional determinants (BLEND) study: design and baseline characteristics. Archives of public health = Archives belges de sante publique. 2016;74:30. doi:10.1186/s13690-016-0140-1.

11. Bernstein L, Ross R. Cancer in Los Angeles County. Los Angeles, CA: University of Southern California. 1991.

12. Tang L, Zirpoli GR, Guru K, Moysich KB, Zhang Y, Ambrosone CB et al. Consumption of raw cruciferous vegetables is inversely associated with bladder cancer risk. Cancer Epidemiology and Prevention Biomarkers. 2008;17(4):938-44. doi:10.1158/1055-9965.EPI-07-2502.

13. Kellen E, Zeegers M, Lousbergh D, Paulussen A, Buntinx F. A Belgian case control study on bladder cancer: rationale and design. Archives of public health = Archives belges de sante publique. 2005;63:17-34.

14. Wakai K, Takashi M, Okamura K, Yuba H, Suzuki K-i, Murase T et al. Foods and nutrients in relation to bladder cancer risk: a case-control study in Aichi Prefecture, Central Japan. Nutrition and cancer. 2000;38(1):13-22. doi:https://doi.org/10.1207/S15327914NC381_3.

15. Lu C-M, Lan S-J, Lee Y-H, Huang J-K, Huang C-H, Hsieh C-C. Tea consumption: fluid intake and bladder cancer risk in Southern Taiwan. Urology. 1999;54(5):823-8. doi:https://doi.org/10.1016/S0090-4295(99)00281-2.

16. Pohlabeln H, Jöckel K-H, Bolm-Audorff U. Non-occupational risk factors for cancer of the lower urinary tract in Germany. European journal of epidemiology. 1999;15(5):411-9. doi:10.1023/a:1007595809278.

17. Steineck G, Hagman U, Gerhardsson M, Norell SE. Vitamin A supplements, fried foods, fat and urothelial cancer. A case-referent study in Stockholm in 1985–87. International journal of cancer. 1990;45(6):1006-11. doi:10.1002/ijc.2910450604.

18. METTLIN C, GRAHAM S. Dietary risk factors in human bladder cancer. American journal of epidemiology. 1979;110(3):255-63. doi:10.1093/oxfordjournals.aje.a112810.

19. Baena AV, Allam MF, Del Castillo AS, Díaz-Molina C, Tapia MJR, Abdel-Rahman AG et al. Urinary bladder cancer risk factors in men: a Spanish case–control study. European journal of cancer prevention. 2006;15(6):498-503. doi:10.1097/01.cej.0000215618.05757.04.

20. Brinkman MT, Karagas MR, Zens MS, Schned A, Reulen RC, Zeegers MP. Minerals and vitamins and the risk of bladder cancer: results from the New Hampshire Study. Cancer Causes & Control. 2010;21(4):609-19. doi:10.1007/s10552-009-9490-0.

21. La Vecchia C, Negri E, Decarli A, Benichou J. Attributable risks for bladder cancer in northern Italy. 1995. doi:10.1016/1047-2797(95)00057-7.

22. Shen M, Hung RJ, Brennan P, Malaveille C, Donato F, Placidi D et al. Polymorphisms of the DNA repair genes XRCC1, XRCC3, XPD, interaction with environmental exposures, and bladder cancer risk in a case-control study in northern Italy. Cancer Epidemiology and Prevention Biomarkers. 2003;12(11):1234-40. doi:PMID: 14652287.

23. Johnson K, Mao Y, Argo J, Dubois S, Semenciw R, Lava J. The National Enhanced Cancer Surveillance System: a case-control approach to environment-related cancer surveillance in Canada. Environmetrics: The official journal of the International Environmetrics Society. 1998;9(5):495-504. doi:https://doi.org/10.1002/(SICI)1099-095X(199809/10)9:5<495::AID-ENV318>3.0.CO;2-H.

24. Ovsiannikov D, Selinski S, Lehmann M-L, Blaszkewicz M, Moormann O, Haenel MW et al. Polymorphic enzymes, urinary bladder cancer risk, and structural change in the local industry. Journal of Toxicology and Environmental Health, Part A. 2012;75(8-10):557-65. doi:https://doi.org/10.1080/15287394.2012.675308.

25. Clavel J, Cordier S. Coffee consumption and bladder cancer risk. International journal of cancer. 1991;47(2):207-12. doi:https://doi.org/10.1002/ijc.2910470208.

26. Hemelt M, Hu Z, Zhong Z, Xie LP, Wong Y, Tam PC et al. Fluid intake and the risk of bladder cancer: Results from the South and East China case-control study on bladder cancer. International journal of cancer. 2010;127(3):638-45. doi:https://doi.org/10.1002/ijc.25084.

27. Cao W, Cai L, Rao JY, Pantuck A, Lu ML, Dalbagni G et al. Tobacco smoking, GSTP1 polymorphism, and bladder carcinoma. Cancer: Interdisciplinary International Journal of the American Cancer Society. 2005;104(11):2400-8. doi:https://doi.org/10.1002/cncr.21446.

28. Taylor JA, Umbach DM, Stephens E, Castranio T, Paulson D, Robertson C et al. The role of N-acetylation polymorphisms in smoking-associated bladder cancer: evidence of a gene-gene-exposure three-way interaction. Cancer research. 1998;58(16):3603-10. doi:PMID: 9721868.

29. van den Brandt PA, Goldbohm RA, van 't Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. Journal of clinical epidemiology. 1990;43(3):285-95. doi:10.1016/0895-4356(90)90009-E.

30. Poortvliet E, Klensin J, Kohlmeier L. Rationale document for the Eurocode 2 food coding system (version 91/2). *Eur J Clin Nutr* 1992;46(Suppl 5):S9-S24.

31. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. New York: Springer; 2009. p. 485-585.

32. Rubin DB. Inference and missing data. Biometrika. 1976;63(3):581-92. doi:https://doi.org/10.1093/biomet/63.3.581.

33. Van Ness PH, Murphy TE, Araujo KL, Pisani MA, Allore HG. The use of missingness screens in clinical epidemiologic research has implications for regression modeling. Journal of clinical epidemiology. 2007;60(12):1239-45. doi:10.1016/j.jclinepi.2007.03.006.

34. Little RJ. A test of missing completely at random for multivariate data with missing values. Journal of the American statistical Association. 1988;83(404):1198-202. doi:http://dx.doi.org/10.1080/01621459.1988.10478722.

35. Li C. Little's test of missing completely at random. The Stata Journal. 2013;13(4):795-809. doi:https://doi.org/10.1177/1536867X1301300407.

36. Pandya R, Pandya J. C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. International Journal of Computer Applications. 2015;117(16):18-21.

37. Quinlan JR. C4. 5: programs for machine learning. Elsevier; 2014.

38. Quinlan JR, editor. Unknown attribute values in induction. Proceedings of the sixth international workshop on Machine learning; 1989: Elsevier.

39. Karaolis M, Moutiris JA, Pattichis CS, editors. Assessment of the risk of coronary heart event based on data mining. 2008 8th IEEE International Conference on BioInformatics and BioEngineering; 2008: IEEE.

40. Lazarou C, Karaolis M, Matalas A-L, Panagiotakos DB. Dietary patterns analysis using data mining method. An application to data from the CYKIDS study. *COMPUT METH PROG BIO* 2012;108(2):706-14. doi:10.1016/j.cmpb.2011.12.011.

41. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. vol 10. Springer series in statistics New York; 2001.

42. Louppe G, Wehenkel L, Sutera A, Geurts P, editors. Understanding variable importances in forests of randomized trees. Advances in neural information processing systems; 2013.

43. De Stefani E, Boffetta P, Ronco AL, Deneo-Pellegrini H, Acosta G, Mendilaharsu M. Dietary patterns and risk of bladder cancer: a factor analysis in Uruguay. Cancer Causes Control. 2008;19(10):1243-9. doi:10.1007/s10552-008-9195-9.

44. Xu C, Zeng XT, Liu TZ, Zhang C, Yang ZH, Li S et al. Fruits and vegetables intake and risk of bladder cancer: a PRISMA-compliant systematic review and dose-response meta-analysis of prospective cohort studies. Medicine (Baltimore). 2015;94(17):e759. doi:10.1097/md.0000000000000759.

45. Vieira AR, Vingeliene S, Chan DS, Aune D, Abar L, Navarro Rosenblatt D et al. Fruits, vegetables, and bladder cancer risk: a systematic review and meta-analysis. *CANCER MED-US* 2015;4(1):136-46. doi:10.1002/cam4.327.

46. Liu H, Wang XC, Hu GH, Guo ZF, Lai P, Xu L et al. Fruit and vegetable consumption and risk of bladder cancer: an updated meta-analysis of observational studies. *EUR J CANCER PREV* 2015;24(6):508-16. doi:10.1097/cej.0000000000000119.

47. Yao B, Yan Y, Ye X, Fang H, Xu H, Liu Y et al. Intake of fruit and vegetables and risk of bladder cancer: a dose-response meta-analysis of observational studies. *CCC* 2014;25(12):1645-58. doi:10.1007/s10552-014-0469-0.

48. Boeing H, Bechthold A, Bub A, Ellinger S, Haller D, Kroke A et al. Critical review: vegetables and fruit in the prevention of chronic diseases. European journal of nutrition. 2012;51(6):637-63. doi:10.1007/s00394-012-0380-y.

49. Riboli E, Norat T. Epidemiologic evidence of the protective effect of fruit and vegetables on cancer risk. The American journal of clinical nutrition. 2003;78(3 Suppl):559s-69s. doi:10.1093/ajcn/78.3.559S.

50. Chatenoud L, Tavani A, La Vecchia C, Jacobs DR, Jr., Negri E, Levi F et al. Whole grain food intake and cancer risk. International journal of cancer. 1998;77(1):24-8. doi:10.1002/(sici)1097-0215(19980703)77:1<24::aid-ijc5>3.0.co;2-1.

51. Augustin LSA, Taborelli M, Montella M, Libra M, La Vecchia C, Tavani A et al. Associations of dietary carbohydrates, glycaemic index and glycaemic load with risk of bladder cancer: a case-control study. *BRIT J NUTR* 2017;118(9):722-9. doi:10.1017/s0007114517002574.

52. Sun JW, Zhao LG, Yang Y, Ma X, Wang YY, Xiang YB. Obesity and risk of bladder cancer: a dose-response meta-analysis of 15 cohort studies. PloS one. 2015;10(3):e0119313. doi:10.1371/journal.pone.0119313.

53. La Vecchia C, Negri E. Nutrition and bladder cancer. Cancer Causes Control. 1996;7(1):95-100. doi:10.1007/bf00115641.

54. Wang C, Jiang H. Meat intake and risk of bladder cancer: a meta-analysis. *MED ONCOL* 2012;29(2):848-55. doi:10.1007/s12032-011-9985-x.

55. Catsburg CE, Gago-Dominguez M, Yuan JM, Castelao JE, Cortessis VK, Pike MC et al. Dietary sources of N-nitroso compounds and bladder cancer risk: findings from the Los Angeles bladder cancer study. *Int J Cancer* 2014;134(1):125-35. doi:10.1002/ijc.28331.

56. Baris D, Waddell R, Beane Freeman LE, Schwenn M, Colt JS, Ayotte JD et al. Elevated Bladder Cancer in Northern New England: The Role of Drinking Water and Arsenic. *JNCI-J NATL CANCER I* 2016;108(9). doi:10.1093/jnci/djw099.

57. Miyata Y, Matsuo T, Araki K, Nakamura Y, Sagara Y, Ohba K et al. Anticancer Effects of Green Tea and the Underlying Molecular Mechanisms in Bladder Cancer. Medicines (Basel). 2018;5(3). doi:10.3390/medicines5030087.

58. Vartolomei MD, Iwata T, Roth B, Kimura S, Mathieu R, Ferro M et al. Impact of alcohol consumption on the risk of developing bladder cancer: a systematic review and meta-analysis. World J Urol. 2019;37(11):2313-24. doi:10.1007/s00345-019-02825-4.

59. Huys R, Jirsa VK. Nonlinear dynamics in human behavior. Springer; 2010.

60. Crutzen R, Giabbanelli P. Using Classifiers to Identify Binge Drinkers Based on Drinking Motives. *SUBST USE MISUSE* 2013;49(1-2):110-5. doi:10.3109/10826084.2013.824467.

61. Rodrigo CP, Aranceta J, Salvador G, Varela-Moreiras G. Food frequency questionnaires. Nutricion hospitalaria. 2015;31(3):49-56.

Supplementary Materials to Chapter 7

Supplementary Table 7-1 Included individual food items

Supplementary Figure 7-1 Receiver Operating Characteristic Curves for Each Cross-validation Run

Supplementary Figure 7-2 Importance Values of Input Variables after C5.0

Supplementary Table 7-1 Included individual food items

Food Items	Food Items	Food Items
Milk and milk products	**Eggs and egg products**	**Meat and meat products**
Milk in miscellaneous	Eggs in miscellaneous	*Meat in miscellaneous*
Liquid milk	Egg dishes	*Beef*
Liquid milk in miscellaneous	Omelette	Beef in miscellaneous
Milk >4% fat		Beef tenderloin
Milk 3-4% fat		Beef Striploin
Milk 1-3% fat		*Veal*
Milk < 1% fat		*Pork*
Processed milk		Pork in miscellaneous
Chocolate-flavoured milk		Pork chump
Acidophilus milk		Pork Loin
Cream		*Mutton/Lamb*
Yoghurt		*Mammals*
Yoghurt in miscellaneous		Horse
Yoghurt >3%		*Poultry*
Yoghurt <1%		Chicken
Cheese		Duck
Cheese in miscellaneous		*Liver*
Soft cheese		*Kidney*
Processed cheese		*Other Offal*
Whey cheese		*Preserved meat*
Imitation milk		Preserved meat in miscellaneous
Imitation milk in miscellaneous		Ham
Non-dairy coffee creamer		Bacon
Milk beverage powder		
Drinking chocolate powder		
Ices		
Dairy ice		
Sorbet		

Supplementary Table 7-1_continued Included individual food items

Food Items	Food Items	Food Items
Fish and fish Products	**Fats and oils**	**Grain and grain products**
Fish in miscellaneous	Fats and oils in miscellaneous	Grain in miscellaneous
Clupeiformes	*Butter*	*Wheat basic products*
Herring	*Margarine*	Oats basic products
Salmon and trout	*Animal fat*	Oatmeal
Whitefish	*Marine Oil*	*Barley basic products*
Perciformes	*Vegetable fats and oil*	Barley meal
Tuna	Vegetable fats and oil in miscellaneous	*Rice basic products*
	Olive oil	Brown rice
	Sunflower oil	Polished rice
	Compound fats and oils	*Basic products of other cereals*
		Pasta and noodles
		Pasta and noodles in miscellaneous
		Leavened bread
		Leavened bread in miscellaneous
		Rye bread
		Wheat bread
		Unleavened bread and crispbreads
		Crispbread
		Bread products
		Fine bakery wares
		Savoury biscuits
		Sweet biscuits and cookies
		Currant bun
		Dough cakes
		Doughnut
		Danish pastry
		Mince pie
		Baked cake

Supplementary Table 7-1_continued Included individual food items

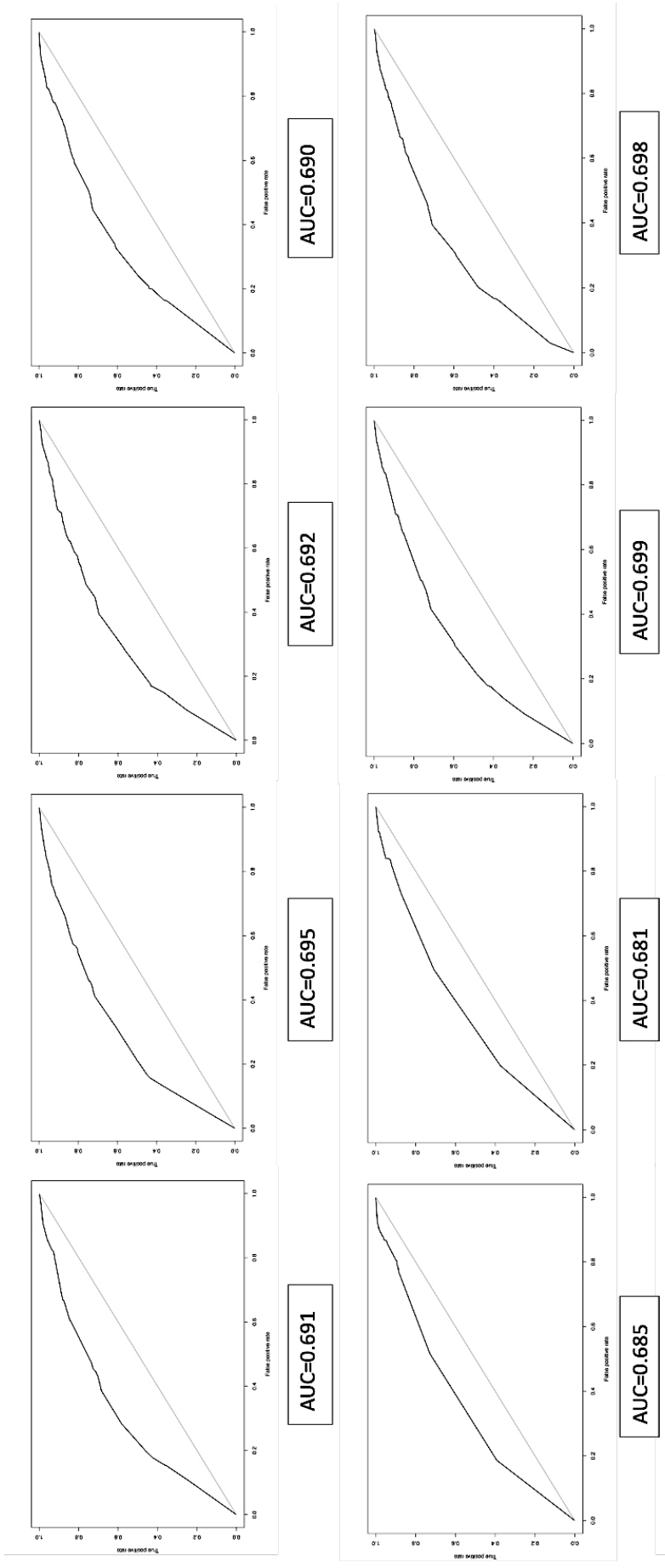
Food Items	Food Items	Food Items
Savoury cereal dishes	**Pulses, seeds, kernels, and nut products**	**Vegetable and vegetable products**
Savoury pancake	Pulses, seeds, kernels, and nut products in miscellaneous	Vegetable in miscellaneous
Risotto	Pulses	*Leaf vegetable*
Savoury pie	Dried broad bean	Endive
Breakfast cereals	Lentil	Lettuce
	Mung bean	Spinach
	Soya bean	*Brassicas*
	Underground pulses	Cauliflower
	Peanut	Broccoli
	Seeds and kernels	Red cabbage
	Sunflower seed	Chinese cabbage
	Pine nut	Cabbage
	Nuts	Turnip tops
	Walnut	Brussels sprout
	Hazelnut	Kohlrabi
	Hickory nut	Curly kale
	Cashew nut	*Stalk vegetables*
	Sweet chestnut	Rhubarb
	Pulse products	Celery
	Unfermented soya paste (tofu)	Fennel
	Peanut butter	*Shoot vegetable*
	Nut and seed products	Asparagus
		Chicory
		Globe artichoke
		Bamboo shoot
		Onion family Vegetable
		Garlic
		Onion
		Leek

Supplementary Table 7-1_continued Included individual food items

Food Items	Food Items	Food Items
Garlic	Runner bean	*Berries*
Chives	Sweet corn	White grapes
Tubers	Okra	Strawberries
Tubers in miscellaneous	Avocado	Blueberries
New potato	Olive	*Citrus fruit*
Main crop	*Edible Fungi*	Orange
Sweet potato	Cultivated mushrooms	Grapefruit
Yam	Truffle	*Miscellaneous*
Root vegetable	Morel	Banana
Carrot	*Seaweed*	Pineapple
Parsnip	*Vegetable mixes*	Kiwi
Swede	Vegetable mixes in miscellaneous	Melon
Beetroot	Mustard and cress	Watermelon
Radish	Pot herb	*Fruit mixture*
Parsley	*Vegetable products*	*Fruit products*
Fruit vegetable	Tomato puree	Fruit products in miscellaneous
Fruit vegetable in miscellaneous	Pickled gherkins	Dried mixed fruit
Tomato	Pickled onion	
Aubergine	Sauerkraut	
Sweet pepper	**Fruit and fruit products**	
Chilli pepper	Fruit in miscellaneous	
Cucumber	*Malaceious fruit*	
Courgette	Apple	
Cucurbita squash	*Prunus species fruit*	
Pod and seed vegetable	Apricot	
Pea	Peach	
Wax beans	Plum	
Broad bean	Sweet cherry	
French bean	*Other stone fruit*	

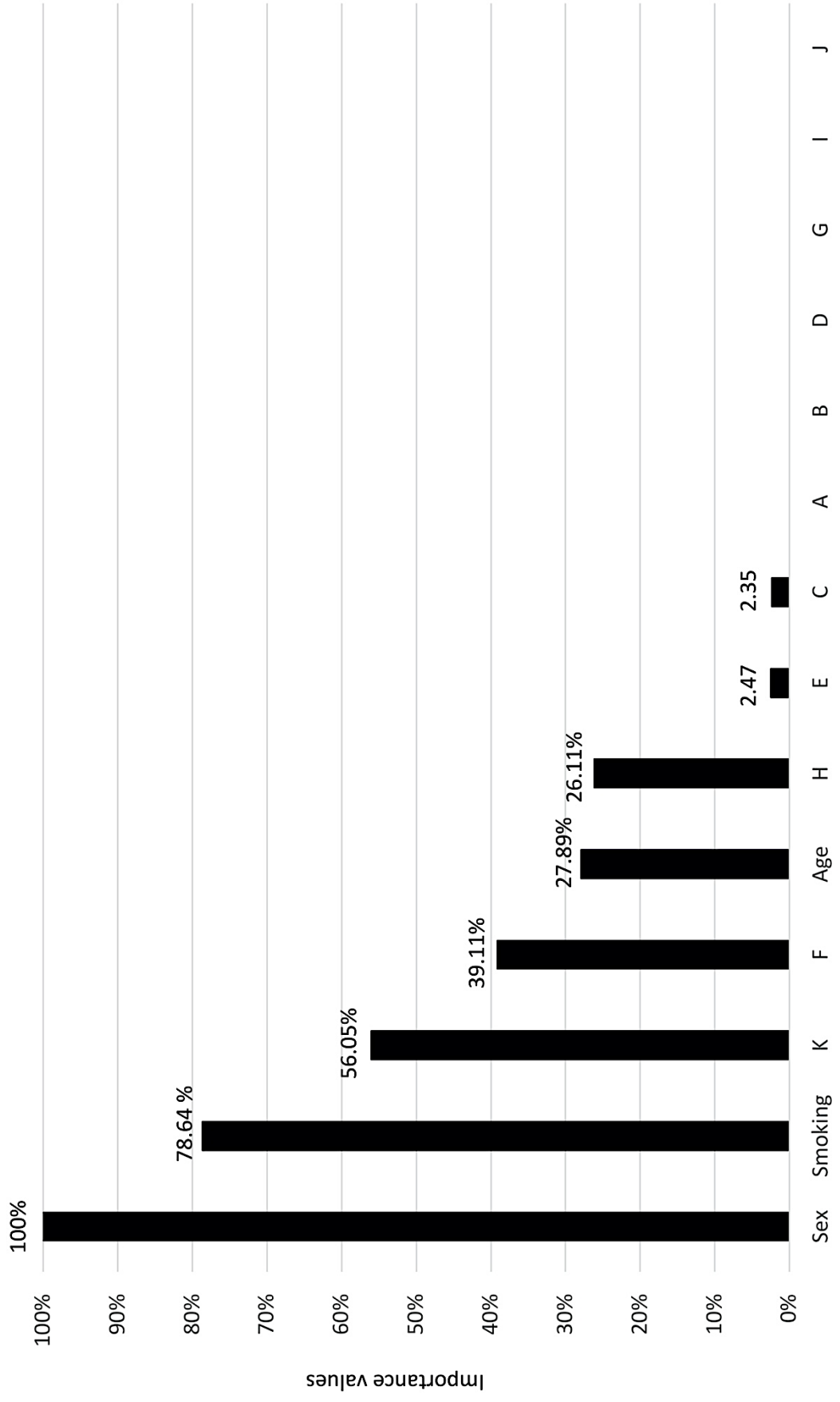
Supplementary Table 7-1_continued Included individual food items

Food Items	Food Items	Food Items
Sugar and sugar products	**Beverages (non-milk)**	Tea in miscellaneous
Sugar in miscellaneous	Beverage in miscellaneous	Herbal tea
Sugar	*Beers in malt*	Decaffeinated coffee
Other sugar	Beer - alcohol >5%	Caffeinated coffee
Fructose	Beer - alcohol 3-5%	Instant coffee
Sugar substitutes	Beer - alcohol 1-3%	Coffee in miscellaneous
Nutritive sweeteners	Beer - alcohol <1%	*Water*
Honey	*Wine*	*Fruit juice*
Syrup	Wine - alcohol >9%	Fruit juice in miscellaneous
Fruit syrup	Wine - alcohol 5-9%	Orange juice
Jams, marmalades and spreads	*Fortified and liqueur*	Grapefruit juice
Chocolate and chocolate products	*Liqueurs*	Apple juice
Chocolate in miscellaneous	*Spirits*	Grape juice
Milk chocolate bar	Whisky	*Vegetable juice*
Chocolate-coated confectionery bars	Gin	Tomato juice
Non-chocolate confectionery	Vodka	*Other juice*
Boiled sweet	*Alcoholic mixed drinks*	
	Punch - alcohol >1%	
	Carbonated soft drinks	
	Carbonated soft drinks in miscellaneous	
	Soda water	
	Cola	
	Lucozade	
	Non-dilution still drinks	
	Still lemonade	
	Infusion drinks	
	Infusion drinks in miscellaneous	
	Black tea	
	Green tea	



Supplementary Figure 7-1 Receiver Operating Characteristic Curves for Each Cross-validation Run

AUC=Area Under the Curve



Supplementary Figure 7-2 Importance Values of Input Variables after C5.0

A: milk and dairy products; B: eggs and egg products; C: meats and meat products; D: fishes and fish products; E: fats, oils and their products; F: grains and grain products; G: pulses, seeds, kernels, nuts and their products; H: vegetables and vegetable products; I: fruits and fruit products; J: sugar and sugar products; K: beverages (non-milk)

The importance values range from 0% to 100%, where 0% indicates 'unimportant' and 100% indicates 'extremely important'.

Chapter 8

Summary and General Discussion

The aim of this thesis was to gain insight in the *health-related quality of life* (HRQoL) of bladder cancer patients at the time of diagnosis and to increase our understanding of the influence of dietary intake on bladder cancer risk by making use of both traditional epidemiology and newer alternatives, *i.e.* machine learning. For these purposes, we investigated; a) the HRQoL in UK bladder cancer patients around time of diagnosis (**Chapter 2**) [based on *Bladder Cancer Patients Prognosis* (BCPP) multi-centre cohort study]; b) the influence of coffee (**Chapter 3-4**), vegetables (**Chapter 5**), grains and dietary fibre (**Chapter 6**), and potential dietary patterns (**Chapter 7**), on bladder cancer risk [based on *BLadder cancer Epidemiology and Nutritional Determinants* (BLEND) international consortium]. This final chapter will summarize and interpret the main findings of this thesis. This will be followed by important methodological and statistical considerations. Finally, recommendations for future research in the field of nutritional epidemiology are given.

8.1 Summary and Interpretation of Main Findings

In **Chapter 2** we showed that patients' HRQoL scores at or around time of bladder cancer diagnosis are comparable to the general population of Germany and possibly the UK. Overall *non-muscle invasive bladder cancer* (NMIBC) patients showed to have a slightly higher HRQoL compared to *muscle invasive bladder cancer* (MIBC) patients, suggesting that progression of bladder cancer might decrease the HRQoL. However, only the domain "Role Functioning" showed to be clinically significantly higher among NMIBC cases. In **Chapter 2** it was also assessed which baseline characteristics at time of diagnosis affect the HRQoL of bladder cancer patients. It was found that females, older age, increased stage, and higher amount of comorbidity negatively affect the HRQoL.

Since bladder cancer can be considered a chronic disease, results of **Chapter 2** might be particularly useful in determining the value of any (drug) therapy and to set baseline values for future intervention studies. Knowing the HRQoL at time of diagnosis enables to measure any HRQoL change caused by the disease and its treatment, which can be used to capture the full effects of disease and its treatment and further discriminate between otherwise similar treatments. This increases awareness among health care professionals to better anticipate on a patients' needs and will directly improves clinical outcomes. In addition, it helps to show

the importance of bladder cancer, thereby creating more opportunities for obtaining new research funding.

While **Chapter 2** investigated the HRQoL of bladder cancer patients at time of diagnosis the remaining chapters of this thesis focussed on the dietary risk factors for bladder cancer. In **Chapter 3** and **Chapter 4** the association between coffee consumption and bladder cancer risk was assessed by pooling data from case-control and cohort studies, respectively. Results from these chapters were conflicting in that results from **Chapter 3** showed an increased bladder cancer risk with higher coffee consumption among never smokers but not smokers, while results from **Chapter 4** showed an increased bladder cancer risk with higher coffee consumption only among male smokers.

It has long been recognized that smoking and coffee consumption are strongly positively associated both observationally (*Swanson et al. 1994* [1]*, Gapstur SM, et al 2017* [2]*, Loftfield E, et al 2017* [3]), as well as genetically (*Sugiyama K et al. 2017* [4]). Therefore, smoking is suggested to be the most important confounder when assessing the association between coffee consumption and disease risk and only completely avoiding residual confounding by smoking could help us in finding the true independent effect of coffee consumption on disease risk. Findings of **Chapter 3** and **Chapter 4** might be seriously influenced by residual confounding of smoking history; in **Chapter 3** the included case-control studies collected data retrospectively from the previous 1-2 year(s), thereby missing essential information on smoking behaviour in early ages which might possibly have affected the bladder cancer risk [5-8]. In contrast, smoking information in **Chapter 4** was only collected at time of study initiation, thereby possibly lacking essential information on start smoking cessation and/or onset of smoking habits. In addition, since smoking is perceived as a socially undesirable behaviour, some individuals, particularly females and adolescents, tend to underreport their smoking behaviours [9-11].

Chapter 5 provides evidence for a protective effect of total vegetable, non-starchy vegetable, and curly kale on bladder cancer risk among females. However, although the nutritional importance of vegetable has long been recognized within the nutritional and medical communities and numerous dietary guidelines suggest higher intake of vegetable as an

important healthy component [12], no protective effect of overall vegetables intake or any other vegetable (sub)types (except for curly kale) on bladder cancer risk could be observed for the whole cohort and neither for males.

Grain intake plays an important role in almost all diets around the world, the health effects of grain intake, and in particular whole grains, have attracted much research interest. However, evidence of their effect on bladder cancer risk is currently mainly lacking. Therefore, in **Chapter 6**, we aimed to explore and compare the effects of whole- and refined grains intake on bladder cancer risk. Here, we found that a higher intake of total whole grain results in a decreased risk of bladder cancer, while no association was observed for the intake of total refined grain. Dietary fibre is the component that distinguishes whole grains from refined grains, therefore, dietary fibre was also assessed in its relation to bladder cancer risk and showed to be inversely associated. These findings support current recommendations to increase the intake of whole grains and dietary fibre to improve public health [13-15].

Since foods and beverages are consumed together, never in isolation of one another, in **Chapter 7** a more holistic dietary approach was taken. Using a machine learning approach, it was aimed to identify and prioritize food groups related to bladder cancer risk. The results of this chapter revealed five food groups (*i.e.* in order of importance: grains and grain products; vegetables and vegetable products; beverages (non-milk); fats, oils and their products; meats and meat products) related to bladder cancer risk (**Table 8-1**).

Table 8-1 *Overview of the Main Methods and Findings of Each Chapter*

Chapters	Study Design	Main Methods	Main Findings
Chapter 2	cohort study	multivariate linear regression model	a) identified a baseline reference value for HRQoL for UK bladder cancer patients around time of diagnosis; b) age, sex, bladder cancer stage and comorbidity were found to significantly influence HRQoL.
Chapter 3	case-control study	multilevel logistic regression model; dose-response analysis	positive associations between coffee consumption and bladder cancer among never smokers but not smokers.

Table 8-1_continued *Overview of the Main Methods and Findings of Each Chapter*

Chapters	Study Design	Main Methods	Main Findings
Chapter 4	cohort study	multilevel Weibull regression model; dose-response analysis	positive associations between coffee consumption and bladder cancer among male smokers but not never smokers and females.
Chapter 5	cohort study	multivariate Cox regression model; dietary diversity score analysis	a) higher intakes of total and non-starchy vegetable are associated with reduced risk of bladder cancer for females but not for males; b) higher intake of curly kale inversely associated with bladder cancer for both sexes.
Chapter 6	cohort study	multivariate Cox regression model; dose-response analysis	higher intakes of total whole grain and total dietary fibre are associated with reduced risk of bladder cancer individually and jointly.
Chapter 7	case-control study and cohort study combined	machine learning (C4.5 algorithm)	identified five key food groups related to bladder cancer risk.

8.2 Methodological Consideration

Study design may influence results from epidemiological studies. Understanding of the strengths and weaknesses of epidemiological studies, and of studies on diet in particular, is essential for their interpretation and for the purposes of setting public policy. Therefore, we will discuss and compare the different study designs used in this thesis and highlight their strengths and limitations.

8.2a Misclassification

8.2a-1 Exposure

In nutritional epidemiological studies, the measurement of diet constitutes a difficult challenge because of the complex nature of the diet itself. Since years ago, methodological issues and particularly issues of questionnaire design have been the subject of research in nutrition. Dietary assessment involves the use of diet/food records, diet-history questionnaires, 24-h recalls, or food-frequency questionnaires (FFQ). Amongst, diet/food records and 24-h recalls are considered most precise, but only useful when measuring short-

term dietary intake. In nutritional epidemiology, however, dietary intake over a longer period is more important, especially when studying the influence of diet on chronic diseases such as bladder cancer. Therefore, FFQs are the most widely used dietary assessment tool in the field of nutritional epidemiology. The use of FFQs, however, is often criticized, since FFQs are subject to considerable measurement errors, resulting from an incomplete listing of all possible foods and from errors in frequency and usual serving size estimates. As a result, the estimated average intake of a study population may be inaccurate and likely biased toward null, *i.e.* misestimation of the true strength of the associations between the dietary exposure in **Chapter 2-7** and bladder cancer. In addition, the workload for a respondent to fill in an FFQ is often considered high in terms of time-consuming, complex and difficult. It is suggested that food intake may be overestimated in relatively long FFQs, whereas shorter lists may underestimate food intake [16]. Since the BLEND consortium includes different studies, each with its own design and questionnaire, the number of items assessed in their specific questionnaire differed significantly, as well as the time needed to complete it. Therefore, there might be differential measurement errors across different studies. However, when comparing mean values and *standard deviations* (SDs) for the intake of total energy, foods and nutrients of interest in **Chapter 3-6**, between long FFQs [17-24] and short FQQs [25-35], by the use of the Spearman's correlation coefficient, all items showed to have a moderate correlation (*i.e.* ≥0.4) (**Table 8-2**) [36], thereby suggesting this misclassification only plays a minor role in the BLEND data.

Table 8-2 *Mean and Standard Deviations for Intake from Long-form FFQs and Short-form FFQs with Spearman's Correlation Coefficient* [a]

Total Energy, Foods and Nutrients Intake	Long-form FFQs	Short-form FFQs	SCC
Total Energy Intake [kcal/day (SD)]	2,084.71 (676.55)	1,953.59 (819.57)	0.67
Coffee Consumption [cups/day (SD)]	3.88 (4.60)	4.12 (3.06)	0.59
Coffee Consumption [ml/day (SD)]	274.33 (319.32)	371.99 (416.63)	0.38

Table 8-2_continued *Mean and Standard Deviations for Intake from Long-form FFQs and Short-form FFQs with Spearman's Correlation Coefficient* [a]

Total Energy, Foods and Nutrients Intake	Long-form FFQs	Short-form FFQs	SCC
Total Vegetable Intake [g/day (SD)]	211.58 (148.24)	138.90 (111.50)	0.45
Total Grain Intake [g/day (SD)]	294.07 (172.28)	132.75 (97.01)	0.47
Total Dietary Fibre Intake [g/day (SD)]	25.01 (12.36)	14.82 (8.30)	0.43

[a] *All correlations were statistically signifiant (P <0.001)*
Abbreviations: SD, standard deviation; FFQ, food frequency questionnaire; SCC, Spearman's correlation coefficient; kcal, kilocalorie; ml, milliliter; g, gram.

8.2a-2 Outcome

Misclassification of bladder cancer could also have introduced bias into our findings, but likely has much less of an impact than misclassification of exposure (*i.e.* dietary intake). Outcomes from the BLEND study rely on the accuracy of methods, used by the included studies, to identify bladder cancer patients and differentiate bladder cancer stages. Although each study ascertained incident bladder cancer with *International Classification of Diseases for Oncology* (ICD-O-3 code C67) using population-based cancer registries, health insurance records, or medical records, misclassification might have occurred especially among small tumours and *carcinoma in situ* (CIS), since the golden standard used in bladder cancer diagnosis (*i.e.* cystoscopy) has poor sensitivity. However, we believe that due to the large number of cases included in BLEND this outcome related misclassification is negligible.

8.2b Causality

All associations between dietary intake and bladder cancer risk observed in this thesis rely on observational data, which is one of the main criticisms levered against nutritional epidemiology and epidemiology in general, since this type of data is deemed to be inferior to experimental data in determining causality. *Randomized clinical trials* (RCTs) are often considered the ultimate approach to provide evidence for causality in nutritional

epidemiology. However, RCTs have some methodological and practical limitations in determining effect of food/nutrient of disease occurring late in life, such as bladder cancer. For example, an RCT can only test a specific intervention, in a specific population, over a relatively short period of time. Therefore, results achieved from RCTs may not yield for a different dose, population or over a longer period of time. Moreover, dietary intervention at a late period in life, which is necessary to ensure an adequate number of bladder cancer cases, might not be effective, since cumulative exposures over time are likely to be important and dietary intake might have the greatest influence early in life. Observational studies overcome these weaknesses, however, two major problems remain using either a case-control or cohort designs to identify causality; a) in both case-control and cohort studies only a limited range of variation in diet can be assessed and b) most cohort studies are not be able to accumulate a sufficient number of cases within a reasonable amount of time. Therefore, combining original data from 27 individual studies conducted in more than 16 different countries and regions across the world, as done for this thesis, is currently considered strongest in determining causality.

Nevertheless, the criteria of causality raised by Bradford Hill (*i.e.* the strength of the evidence, consistency across populations, specificity, temporality, dose-response and biological plausibility) should be carefully taken into account when determining the weight of the evidence. Regarding the first criterion (*i.e.* the strength of the evidence) the associations described in this thesis are not as high as Bradford Hill's initial recommendations and examples to sustain this point in proving causality [37]. However, the weak associations found in this thesis do not necessarily imply lack of causality, since strict causal criteria in current practice, especially for judging weak associations, may diminish the validity of causal inferences and produce ungeneralizable results [38-41]. Therefore, the observed weak association in this thesis may still have large public health effects, especially knowing the dietary factor assessed are common and bladder cancer presents an important public health concern. In addition, even if causality is questionable, general recommendation regarding dietary intake should be made on the best evidence available, in order to help individuals to make healthy food choices every day. However, the causal interpretation of the associations observed in thesis between coffee consumption and bladder cancer risk (**Chapter 3** and

Chapter 4), which differ among different study designs, should be taken with caution, since these are most likely biased by residual confounding of smoking.

Results from both case-control and cohort studies may also have been affected by reverse causality. In cohort studies bladder cancer cases may have experienced symptoms from their disease before diagnosis, which may have caused individuals to change their dietary behaviour. Their reported dietary behaviour would, therefore, not be represent their long-term dietary behaviour, thereby affecting the risk estimates. Especially results from **chapter 4** could have been affected by this type of bias, since early symptoms of bladder cancer included as frequent urination and sudden urges, which may have caused individuals to consume less coffee to experience less urinary problems that may have occurred. To control for this, we performed sensitivity analysis excluding cases diagnosed within the first two years of follow-up. Results, however, did not change. In case-control studies, a similar situation could have occurred, in which the time period covered by the FFQs is too short capturing dietary behaviour before they changed as a consequence of bladder cancer symptoms. Unlike in cohort studies, case-control studies do not allow to control for this reverse causality. Results from **chapter 3** might therefore been affected by this.

8.2c Bias

Bias is defined as any tendency which prevents unprejudiced consideration of a question. A thorough understanding of bias and how it affects study results is essential for making valid conclusions. Therefore, in this section we describe all potential biases that may have occurred during the conduction of studies described in this thesis.

The BCPP cohort study described in **Chapter 2**, used semi-structured face-to-face interviews by a trained research nurse to collect data on HRQoL. However, due to practical reasons, the research nurse was unable to interview all patient face-to-face, and data was collected by the use of semi-structured telephone interview. Telephone interviews do not allow the researchers to draw from the visual aspects such as facial and body expression and the ability to gauge the interviews. Therefore, the data collection in this study might have introduced bias. However, no statistical differences in results were found when comparing the face-to-face data to the data collected by the use of a telephone.

Finally, the low response rate observed in the BCPP study could have introduced bias, by means of a lower response rate among bladder cancer cases with inferior prognosis. Comparing baseline characteristics of responders and non-responders, however, showed no major difference, indicating non-response is random and therefore unlikely to introduce bias.

Since the BLEND study (**Chapter 2-7**) pooled data from individual studies, bias may have been introduced in both the data collection method used by the individual study level as well as in the pooling process. On study level, bias might have been introduced by the use of a self-reported data collection method, which may arise from social desirability, recall period, and/or selective recall. Social desirability in studies assessing the influence of diet on disease risk, is the tendency of study participants to deny their unhealthy food consuming and ascribe to themselves healthy dietary habits that are socially desirable. This may lead to overreporting "good/healthy" food items, or underreporting "bad/unhealthy" food items, thereby affect the risk estimates. In addition, as mentioned previously, the most important risk factor for bladder cancer, such as smoking, is also likely biased by social desirability. This type of bias, may especially have affected results of **Chapter 5**, since vegetables are well-known for being good for health, and they are biologically interacted with smoking [42].

Recall bias is another type of self-reported bias, which mainly may have occurred in the BLEND included case-control studies, where participants were required to evaluate their dietary intake retrospectively. This, particularly, becomes problematic if cases, due to experiences of a life-changing event (*i.e.* bladder cancer diagnosis) may have a better memory of their dietary exposure compared to controls. However, although this issue has been addressed and analysed for its consequences in many epidemiological/methodological papers [43-46], no clear answer on the magnitude of the effect of this specific type of bias could be drawn. In addition, concerns are especially expected for those dietary exposures that attracted public attention (*e.g.* fruit and vegetable), since the knowledge of an unhealthy or healthy effect of such an exposure is a prerequisite for study subjects to change their reporting behaviour of individual exposure. In **Chapter 3** data from case-control studies was used to describes the relation between coffee consumption and bladder cancer risk. Since coffee is a daily-consumed beverage of which, so far, no clear evidence of a positive nor negative effect on

health has been reported, the magnitude of recall bias in terms of coffee consumption is likely to be minimal. In **Chapter 7**, data from case-control studies was combined with data from a nested case-cohort study, therefore, the role of recall bias is likely to be diminished. In addition, tenfold cross-validation method applied int this chapter randomly divided the cohort and case-control data into 10 parts, therefore recall bias coming from the included case-control studies was diminished [47]. Furthermore, C5.0 classification algorithm aimed to find the balanced trade-off of high or low intake of each food group based on whole population, therefore we assumed the effect of individual recall bias was minimised, particularly regarding the relatively small proportion of participants from case-control studies.

In prospective cohort studies, problems of recall are avoided by asking for information at the time that participants joined the study. Therefore, results of **Chapter 4-6** are not affected by this type of bias.

Besides self-reported bias, selection bias is another form of bias that might have been introduced on study level. In the BLEND included case-control studies, selection bias might have occurred by the selection of a control group not representative of the population that gave rise to cases and/or the exposure factor of interest was influenced by an underlying condition other than bladder cancer. The latter might especially be the case in hospital-based control selection. However, in **Chapter 3** stratified analysis on control selection did not result in differences in risk estimates, thereby suggesting no risk estimates inflated or hampered by the selection of controls. In addition, the likelihood this bias was introduced in the included case-control studies was reduced by the fact that several studies conducted matching case-control methodology at the beginning of participants recruitment [17, 21, 22, 25-27, 30, 31, 48].

In the BLEND included cohort studies selection bias might have been introduced if the rates of participation or the rates of loss to follow-up differ by both the dietary exposure of interest and/or bladder cancer status. This especially applies to cohort studies who sampled their participants in health-related practices [32, 35]. Individuals attending health-related practices are more likely health conscious than the average population, thereby having a healthier dietary behaviour, resulting in an apparently affected association between the exposure of

interest and bladder cancer. However, most included cohort studies were reported to have a representative population for residents in area of study conduction.

Bias in cohort studies may also arise by missing outcome information due to death. It could have occurred that bladder cancer was not captured before study participants died from any cause other than bladder cancer, leading to an underestimation of bladder cancer incidence and a risk set no longer representable of the sample. This bias will, therefore, also affect the risk estimates. However, since bladder cancer is mainly diagnosed at an early disease stage, it is unlikely that being an undiagnosed bladder cancer case increases the risk of dying compared to non-cases. Therefore, this missing outcome information bias is expected to play (if any) a minor role in analysis derived from BLEND. In addition, in cohort studies where survival analyses were applied, as described in **Chapter 4-6**, competing risks might have caused a significant problem. Competing risks could have occurred when study participants experienced one or more events or outcomes which "compete" with bladder cancer. In those individuals, the competing risk hinders the observation of bladder cancer or modifies the chance of bladder cancer to occur. Unfortunately, in BLEND no specific information on loss-to-follow-up participants or death cause was available, and therefore, no competing-risk analysis could be performed in order to assess the magnitude of this type of bias.

8.2d Pooled data

Although pooled data analyses have major advantages, including; a) the ability to perform detailed and subgroup analysis with enough statistical power, b) the potential to detect small effect sizes, c) capturing a wide variety of dietary items, and d) standardization of covariates, the pooling process may have introduced errors. The food classification system used in this thesis (*i.e.* Eurocode 2 Core classification version 99/2 [49]) is especially designed and relevant for food consumption data of European studies. It might, therefore, be possible that some ethnic foods important in certain geographical areas outside of Europe have distinct nutritional composition and were not covered by the classification system used in this thesis. This may have led to inaccurate translation of the dietary data from studies originating from outside of Europe. In addition, the pooling process added layers of complexity in terms of missing data and residual confounding, which may vary significantly across the individual studies.

8.2e Statistical validity including implications of confounder selection

In this thesis, several types of traditional statistical models, *e.g.* logistic regression model, Weibull regression model and Cox regression model, were applied as well as more advanced statistical models, *e.g.* machine learning.

Since analysis in the BLEND database were based on large-scale data from multiple studies conducted in different countries, we used multilevel (*i.e.* level-1: individual participant, level-2: case-control study centre) logistic regression model in **Chapter 3** [50], multilevel (*i.e.* level 1: individual participant, level-2: cohort study centre) Weibull regression model in **Chapter 4** [51]) and Cox regression model stratified on cohort in **Chapter 5** and **Chapter 6** [52]. The multilevel logistic regression model used in **Chapter 3**, in which the individual (*i.e.* level-1) were nested in the study centre (*i.e.* level-2), takes into account the odds that the outcome variable equals one (rather than zero) may vary from one study centre to another (*i.e.* the intercept may vary) and the effect of a lower-level variable (*i.e.* individual participant) may also vary from one cluster to another (*i.e.* the slope may vary), thereby adjusting for across study centre heterogeneity.

While the semi parametric-Cox (proportional hazard) regression model is the most commonly used model for survival analysis, we chose to use the parametric Weibull regression model to analyse the data in **Chapter 4**. Both models allow for the inclusion of multiple covariates including the survival time, assume proportional hazards and both provide asymptotically unbiased, equally efficient estimates of the hazard ratio. However, the Cox model has less restrictive assumptions. If both models are applicable, as in **Chapter 4**, the parametric Weibull model may offer advantages over the Cox model in terms of accuracy and efficiency. In **Chapter 5** and **Chapter 6**, however, the Weibull distribution was violated, and was therefore replaced by the Cox model.

For the dose-response analysis in this thesis, *restricted cubic spline* regression (RCS) model (**Chapter 4**) [53] or polynomial regression model (**Chapter 3, Chapter 5** and **Chapter 6**) [54, 55] were use. Both models are powerful to characterize a dose-response association between a continuous exposure and an outcome, to visually and/or statistically check the assumption

of linearity of the association, and to minimize residual confounding when adjusting for another continuous exposure. For the RCS model, knots at the 5th, 35th, 65th and 95th percentile of the exposure distribution were chosen, which resulted in categories with reasonably large numbers of both bladder cancer cases and non-cases in each category. Results from these dose-response analyses were of specific added valuable to define a threshold of the observed effect of the dietary intake of interest on bladder cancer risk.

A machine learning approach was systematically and successfully applied in **Chapter 7**. Machine learning is an innovative approach in the field of nutritional epidemiology, and has demonstrated its potential to complement traditional statistical analysis, particularly for nonlinear phenomena such as our dietary habits [56], and without requiring a priori assumptions on the relationship between diet and health outcomes [57]. However, machine learning cannot be used to investigate the causal relation and might particularly be useful for hypothesis generation rather than hypothesis testing. Therefore, the effects of the food-groups identified in **Chapter 7** on bladder cancer needs further investigation.

In all analyses throughout this thesis, selection and handling of confounders plays an essential role. In this thesis confounding could have arose when a factor is associated with both the dietary factor under investigation (**Chapter 3-6**) and bladder cancer, and is not part of the causal pathway between these two factors. Ideally, confounding should already be controlled for at the design phase of the study by restriction of the selected study population. However, since BLEND gathered data from previously conducted studies, no influence could be applied on the design of the studies. Therefore, adjustment for confounding was a crucial step during the analysis phase. All analysis included the most important risk factors (*i.e.* age, sex and smoking) for the development of bladder cancer; in addition, other potential risk factors identified via literature and textbook were tested for their potential influence on the bladder cancer risk estimates. Although identified as potential risk factors for bladder cancer, no information was available on physical activity and *body mass index* (BMI); however, as a proxy measure we used total energy intake. Since all main risk estimates of this thesis remained stable after adjustment for all potential and available confounders, it is unlikely that inappropriate adjustments have markedly affected the findings of this thesis. However, residual confounding by unknown factors could still have influenced our results.

8.2f Multiple Testing

Another concern that needs to be considered is whether the results were caused by chance alone. Within this thesis more than 10 tests were performed, it is therefore, possible that the findings are a consequence of multiple testing. However, although large numbers of statistical techniques are available to correct for multiple comparison, it still remains a topic of discussion whether and how to correctly adjust for multiple testing [58, 59]. Given the fact that most of the observed significant findings (*i.e.* total vegetable, non-starchy vegetable, total whole grain and total dietary fibre) are supported by a plausible biological evidence, it is unlike these results are chance findings.

8.3 Future Aspects of Cancer- and Nutritional Research

During the last decade, the field of nutritional epidemiology has evolved from a small speciality towards a major contributor to science. The field advanced significantly in the last 10 years, new approaches gained ground and helped to inform dietary guidelines and recommendations. In the near future, a paradigm shift in how to establish nutritional epidemiological research is expected to happen. Therefore, this section will describe the new, promising, developments in the basis of nutritional epidemiology.

8.3a Dietary Assessment

As previously mentioned, FFQs are the prime assessment instrument in nearly all long-term studies. The use of this, however, is still facing a lot of criticism, and validation studies of FFQs with a reference instrument and/or biomarker revealed that the measurement error of FFQs is significant [60, 61], thereby distorting the risk estimates and leading to inaccurate conclusions. Therefore, there is high need for improvement of the quality of the dietary data. Here, several options will be discussed.

8.3a-1 Biomarkers

The inclusion of biomarkers could potentially help in reducing the error rate in dietary data. Biomarkers are able to objectively assess dietary intake without the bias of self-reported errors and overcome the problem of interaction effect between food items/nutrients. So far, some dietary biomarkers, present in blood, urine, adipose tissue, or hair, have been identified

as promising markers for dietary intake [62, 63]. These include biomarkers that could potentially have increased the accuracy of the risk estimates established in this thesis, namely; a) two caffeine metabolites, *i.e.* 17U and 1X, which can be used as biomarkers for caffeine intake [64]; b) serum carotenoids, plasma vitamin C, urinary potassium, flavonoids in both urine and serum, possessing the ability to assess the intake of fruit and vegetables [65, 66]; and c) plasma *alkyl resorcinol* (AR) concentrations, which have been suggested to be able to accurately measure whole grain intake [67, 68].

Although biomarkers are very promising in terms of accurate measurement of dietary intake, one should be aware of the limitations and practical challenges. Factors such as genetic variability, lifestyle factors that might interact with the nutrient under investigation, (*e.g.* smoking), the type of biological sample used, and the analytical methodology may skew the dietary measurement of biomarkers. In addition, the costs and invasiveness of biomarkers might lead to reduced numbers of individuals willing to participate. More research is needed to address these limitations, and to develop cost-effective and non-invasive techniques.

8.3a-2 Omics

Similar dietary patterns may cause different health effects in different populations. A key aim of nutritional research therefore is to determine the role of diet in metabolic regulation and to identify factors that influences an individual's response to diet. Multiple new opportunities regarding this aim have recently been created by remarkable advances in omics technologies, including genomics, metabolomics and proteomics. These novel omics technologies hold great potential to get deeper insights in the complex relation between food/nutrient and metabolism.

Genomics is the study of all of a person's genes (the genome), including interactions of those genes with each other and with the person's environment. Nutrigenomics, therefore, defines the interaction between nutrition and individual's genome by *Davis and Hord, 2005* [69]. The field of nutrigenomics has already increased our understanding of the enormous complexity of human carcinogenesis affected by dietary intake [69, 70]. Multiple genetic variations in nutrition-metabolizing enzyme systems have been demonstrated to modify the effect of dietary exposures on cancer risk [71]. For example, several studies showed that the effect of

caffeine on overall health might be more harmful in individuals with a genetic predisposition towards being slow caffeine metabolizers [72]. Since results from **Chapter 4** indicate that the harmful effect of coffee might be caused by the caffeine, slow caffeine metabolizers might be particularly at risk for the development of bladder cancer when consuming high amounts of coffee. This clearly shows that future research in the field of nutritional epidemiology will greatly benefit from this relatively novel nutrigenomic area, which can be seen as a conceptual analogue to personalized medicine, the concept of adapting food to individual need.

Metabolomics is the study of metabolites that derived from exogenous sources (*e.g.* diet, medications, environmental exposures) or produced endogenously as a result of metabolic processes [73]. As a relatively new field of metabolomics, "nutritional metabolomics" is the study of endogenous and gut microbiota metabolic response to dietary intake and the identification of metabolites that originate from food and could be used as biomarkers of exposure of these foods [74]. It thereby, offers the great potential to help us discovering new biomarker discovery, to map the complex metabolism of food consumption and the unravel the biological consequences of different diets [75-78].

Proteomics is concerned with the systematic, high-throughput approach to protein expression analysis of a cell or an organism. Proteins are the final and active product of most of the genome and thus, their levels are the most accurate reflection of what is happening when gene expression is modulated by food intake. It therefore, offers an unique opportunity to unravel the effects of dietary components, on protein networks that impinge on cancer [79]. In nutrition, the field of proteomics already identified multiple molecular evidence for health benefits of several dietary items [80-87], however, the field is still in its infancy and needs further examination.

8.3b Lifetime exposure

There is growing evidence that chronic disease (*e.g.* cancer) risk may already begin in early life and that nutritional exposure during this period may have long lasting consequences. Therefore, a life course approach to nutritional epidemiology, in which long-term effects of dietary exposures during gestation, childhood, adolescence, young adulthood, and later adult

life on later disease risk [88] are investigated, is currently receiving increased attention. Evidence, however, is still scarce, especially in terms of cancer, which might due to a) investigating life course processes for cancer requires measuring data at multiple time points throughout the life span; b) collection of adequate exposure measures, such as biomarker levels, at multiple points is quite expensive to perform. Since the importance of the life course as a framework in cancer research is increasingly being recognized, future research is needed to examine diets at various times in life and with long follow-up.

8.3c Novel and advanced analysis approach

With the rise of the omics data, nutritional epidemiology is entering the "big data" era, as such, the potential for advanced analysis grows significantly. The use of traditional regression and/or survival analysis will likely be replaced by machine learning techniques. Particularly in contexts where prediction or hypothesis generation rather than hypothesis testing is the analytic goal. Machine learning technique applied in the field of nutritional epidemiology could help to identifying dietary patterns that relate variables to diseases and maximize accuracy when predicting those outcomes. In addition, algorithmic identification of dietary patterns/items associated with a disease of interest allows nutritional epidemiologists to focus on independent validation and interpretation of these associations in subsequent studies. Although this relative new approach of hypothesis generation is yet in a very early stage, results from **chapter 7** proves this technique might potentially reveal novel dietary leads.

Finally, the omics data from different fields will enable nutritional epidemiologists to move from identifying individual dietary items/patterns to analysing more complex notions: gene polymorphisms (nutrigenetics), DNA methylation (nutritional epigenomics), gene expression (nutritional transcriptomics), altered formation or bioactivation of proteins (proteomics), and characterizing how the quantity and timing of exposure influence small molecular weight cellular constituents (metabolomics). However, although the use of big data in nutritional epidemiology calls for new skills and collaboration between different research field (*i.e.* knowledge engineering), traditional epidemiological and expert skills will not be removed due to their bona fide multidisciplinary features in understanding how political, social and scientific factors intersect to disease risk.

In summary, these future aspects and novel techniques for research of diet and nutrition in cancer prevention requires infrastructure development and solid proof of principle; however, they could considerably increase knowledge of the relationship between diet and cancer.

8.4 Conclusion

In summary, there is no doubt that over the last decades nutritional epidemiology has provided clear evidence that dietary intake plays an essential role in the development of diseases, including bladder cancer, and that this knowledge has been successfully translated into dietary guidelines. This thesis suggests, for preventing bladder cancer by nutritional modification, that one should avoid a daily coffee consumption of more than 4 cups, consume high amounts of non-starchy vegetables (>200 g/day), and consume high amounts of whole grains (>15g/day) and dietary fibre (>25 g/day). In addition, this thesis identified several key food groups [*e.g.* grains and grain products; vegetables and vegetable products; beverages (non-milk); fats, oils and their products; meats and meat products] that may have an essential effect on bladder cancer development.

We believe that by incorporating a multidisciplinary approach applied across the life cycle, the field of nutritional epidemiology holds great promise in terms of disease prevention, including the prevention of bladder cancer.

References

1. Swanson JA, Lee JW, Hopp JW. Caffeine and nicotine: a review of their joint use and possible interactive effects in tobacco withdrawal. Addict Behav. 1994;19(3):229-56. doi:10.1016/0306-4603(94)90027-2.

2. Gapstur SM, Anderson RL, Campbell PT, Jacobs EJ, Hartman TJ, Hildebrand JS et al. Associations of Coffee Drinking and Cancer Mortality in the Cancer Prevention Study-II. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2017;26(10):1477-86. doi:10.1158/1055-9965.Epi-17-0353.

3. Loftfield E, Freedman ND, Inoue-Choi M, Graubard BI, Sinha R. A Prospective Investigation of Coffee Drinking and Bladder Cancer Incidence in the United States. Epidemiology (Cambridge, Mass). 2017;28(5):685-93. doi:10.1097/ede.0000000000000676.

4. Sugiyama K, Sugawara Y, Tomata Y, Nishino Y, Fukao A, Tsuji I. The association between coffee consumption and bladder cancer incidence in a pooled analysis of the Miyagi Cohort Study and Ohsaki Cohort Study. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP). 2017;26(2):125-30. doi:10.1097/cej.0000000000000226.

5. Rink M, Furberg H, Zabor EC, Xylinas E, Babjuk M, Pycha A et al. Impact of smoking and smoking cessation on oncologic outcomes in primary non–muscle-invasive bladder cancer. European urology. 2013;63(4):724-32.

6. Rink M, Zabor EC, Furberg H, Xylinas E, Ehdaie B, Novara G et al. Impact of smoking and smoking cessation on outcomes in bladder cancer patients treated with radical cystectomy. European urology. 2013;64(3):456-64.

7. Chen CH, Shun CT, Huang KH, Huang CY, Tsai YC, Yu HJ et al. Stopping smoking might reduce tumour recurrence in nonmuscle-invasive bladder cancer. BJU international. 2007;100(2):281-6.

8. Grotenhuis AJ, Ebben CW, Aben KK, Witjes JA, Vrieling A, Vermeulen SH et al., editors. The effect of smoking and timing of smoking cessation on clinical outcome in non–muscle-invasive bladder cancer. Urologic Oncology: Seminars and Original Investigations; 2015: Elsevier.

9. Ward J, Sanson-Fisher R. Accuracy of patient recall of opportunistic smoking cessation advice in general practice. Tobacco Control. 1996;5(2):110-3.

10. Connor Gorber S, Schofield-Hurwitz S, Hardt J, Levasseur G, Tremblay M. The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status. Nicotine Tob Res. 2009;11(1):12-24. doi:10.1093/ntr/ntn010.

11. Wong SL, Shields M, Leatherdale S, Malaison E, Hammond D. Assessment of validity of self-reported smoking status. Health reports. 2012;23(1):47-53.

12. Kushi LH, Doyle C, McCullough M, Rock CL, Demark-Wahnefried W, Bandera EV et al. American Cancer Society Guidelines on nutrition and physical activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity. CA Cancer J Clin. 2012;62(1):30-67. doi:10.3322/caac.20140.

13. European Food Information Council (EUFIC). https://wwweuficorg/en.Access July 15, 2019.

14. US Department of Agriculture and US Department of Health and Human Services. Dietary guidelines for Americans. https://wwwchoosemyplategov/dietary-guidelines.Accessed July 15, 2019.

15. The European Commission's science and knowledge service. Available at https://eceuropaeu/jrc/en/health-knowledge-gateway/promotion-prevention/nutrition/fibre.Accessed September 17, 2019.

16. Boushey CJ, Coulston AM, Rock CL, Monsen E. Nutrition in the Prevention and Treatment of Disease. Elsevier; 2001.

17. Jiang X, Castelao JE, Groshen S, Cortessis VK, Ross RK, Conti DV et al. Alcohol consumption and risk of bladder cancer in Los Angeles County. International journal of cancer. 2007;121(4):839-45. doi:10.1002/ijc.22743.

18. Kellen E, Zeegers M, Paulussen A, Van Dongen M, Buntinx F. Fruit consumption reduces the effect of smoking on bladder cancer risk. The Belgian case control study on bladder cancer. International journal of cancer. 2006;118(10):2572-8. doi:10.1002/ijc.21714.

19. Steineck G, Hagman U, Gerhardsson M, Norell SE. Vitamin A supplements, fried foods, fat and urothelial cancer. A case-referent study in Stockholm in 1985-87. International journal of cancer. 1990;45(6):1006-11. doi:10.1002/ijc.2910450604.

20. Johnson K, Mao Y, Argo J, Dubois S, Semenciw R, Lava J. The National Enhanced Cancer Surveillance System: a case-control approach to environment-related cancer surveillance in Canada. Environmetrics: The official journal of the International Environmetrics Society.

1998;9(5):495-504. doi:https://doi.org/10.1002/(SICI)1099-095X(199809/10)9:5<495::AID-ENV318>3.0.CO;2-H.

21. Cao W, Cai L, Rao JY, Pantuck A, Lu ML, Dalbagni G et al. Tobacco smoking, GSTP1 polymorphism, and bladder carcinoma. Cancer. 2005;104(11):2400-8. doi:10.1002/cncr.21446.

22. D'Avanzo B, La Vecchia C, Negri E, Decarli A, Benichou J. Attributable risks for bladder cancer in northern Italy. Annals of epidemiology. 1995;5(6):427-31.

23. van den Brandt PA, Goldbohm RA, van 't Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. Journal of clinical epidemiology. 1990;43(3):285-95. doi:10.1016/0895-4356(90)90009-E.

24. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol. 1997;26 Suppl 1:S6-14. doi:10.1093/ije/26.suppl_1.s6.

25. Pohlabeln H, Jockel KH, Bolm-Audorff U. Non-occupational risk factors for cancer of the lower urinary tract in Germany. European journal of epidemiology. 1999;15(5):411-9.

26. Clavel J, Cordier S. Coffee consumption and bladder cancer risk. International journal of cancer. 1991;47(2):207-12. doi:10.1002/ijc.2910470208.

27. Taylor JA, Umbach DM, Stephens E, Castranio T, Paulson D, Robertson C et al. The role of N-acetylation polymorphisms in smoking-associated bladder cancer: evidence of a gene-gene-exposure three-way interaction. Cancer Res. 1998;58(16):3603-10. doi:PMID: 9721868.

28. Lu CM, Lan SJ, Lee YH, Huang JK, Huang CH, Hsieh CC. Tea consumption: fluid intake and bladder cancer risk in Southern Taiwan. Urology. 1999;54(5):823-8. doi:10.1016/s0090-4295(99)00281-2.

29. Baena AV, Allam MF, Del Castillo AS, Diaz-Molina C, Requena Tapia MJ, Abdel-Rahman AG et al. Urinary bladder cancer risk factors in men: a Spanish case-control study. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP). 2006;15(6):498-503. doi:10.1097/01.cej.0000215618.05757.04.

30. Shen M, Hung RJ, Brennan P, Malaveille C, Donato F, Placidi D et al. Polymorphisms of the DNA repair genes XRCC1, XRCC3, XPD, interaction with environmental exposures, and bladder cancer risk in a case-control study in northern Italy. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2003;12(11 Pt 1):1234-40.

31. Mettlin C, Graham S. Dietary risk factors in human bladder cancer. Am J Epidemiol. 1979;110(3):255-63. doi:10.1093/oxfordjournals.aje.a112810.

32. Ozasa K, Shimizu Y, Sakata R, Sugiyama H, Grant EJ, Soda M et al. Risk of cancer and non-cancer diseases in the atomic bomb survivors. Radiation protection dosimetry. 2011;146(1-3):272-5. doi:10.1093/rpd/ncr168.

33. Giles GG, English DR. The Melbourne Collaborative Cohort Study. IARC scientific publications. 2002;156:69-70. doi:PMID: 12484128.

34. Milne RL, Fletcher AS, MacInnis RJ, Hodge AM, Hopkins AH, Bassett JK et al. Cohort Profile: The Melbourne Collaborative Cohort Study (Health 2020). Int J Epidemiol. 2017;46(6):1757-i. doi:10.1093/ije/dyx085.

35. White E, Patterson RE, Kristal AR, Thornquist M, King I, Shattuck AL et al. VITamins And Lifestyle cohort study: study design and characteristics of supplement users. Am J Epidemiol. 2004;159(1):83-93. doi:10.1093/aje/kwh010.

36. Nelson M. The validation of dietary assessment. Design concepts in nutritional epidemiology. 1997;2:241-56.

37. Hill AB. The environment and disease: association or causation? : Sage Publications; 1965.

38. Charlton BG. Attribution of causation in epidemiology: chain or mosaic? Journal of clinical epidemiology. 1996;49(1):105-7. doi:10.1016/0895-4356(95)00030-5.

39. Potischman N, Weed DL. Causal criteria in nutritional epidemiology. The American journal of clinical nutrition. 1999;69(6):1309s-14s. doi:10.1093/ajcn/69.6.1309S.

40. Weed DL, Gorelic LS. The practice of causal inference in cancer epidemiology. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 1996;5(4):303-11.

41. Weed D. Causal and preventive inference. Cancer prevention and control. 1995;285:302.

42. Tang L, Zirpoli GR, Jayaprakash V, Reid ME, McCann SE, Nwogu CE et al. Cruciferous vegetable intake is inversely associated with lung cancer risk among smokers: a case-control study. BMC cancer. 2010;10:162. doi:10.1186/1471-2407-10-162.

43. Barry D. Differential recall bias and spurious associations in case/control studies. Stat Med. 1996;15(23):2603-16. doi:10.1002/(sici)1097-0258(19961215)15:23<2603::aid-sim371>3.0.co;2-g.

44. Barry D, Livingstone V. The investigation and correction of recall bias for an ordinal response in a case-control study. Stat Med. 2006;25(6):965-75. doi:10.1002/sim.2238.

45. Chouinard E, Walter S. Recall bias in case-control studies: an empirical analysis and theoretical framework. Journal of clinical epidemiology. 1995;48(2):245-54.

46. Gefeller O. Invited commentary: Recall bias in melanoma -- much ado about almost nothing? Am J Epidemiol. 2009;169(3):267-70; discussion 71-2. doi:10.1093/aje/kwn362.

47. Vanwinckelen G, Blockeel H, editors. On estimating model accuracy with repeated cross-validation. Benelearn 2012: Proceedings of the 21st belgian-dutch conference on machine learning; 2012.

48. Cole JA, Taylor JS, Hangartner TN, Weinreb NJ, Mistry PK, Khan A. Reducing selection bias in case-control studies from rare disease registries. Orphanet journal of rare diseases. 2011;6(1):61.

49. Poortvliet E, Klensin J, Kohlmeier L. Rationale document for the Eurocode 2 food coding system (version 91/2). *Eur J Clin Nutr* 1992;46(Suppl 5):S9-S24.

50. Wong GY, Mason WM. The hierarchical logistic regression model for multilevel analysis. Journal of the American Statistical Association. 1985;80(391):513-24.

51. Zhang Z. Parametric regression model for survival data: Weibull regression model as an example. Annals of translational medicine. 2016;4(24).

52. Giganti MJ, Luz PM, Caro-Vega Y, Cesar C, Padgett D, Koenig S et al. A Comparison of Seven Cox Regression-Based Models to Account for Heterogeneity Across Multiple HIV Treatment Cohorts in Latin America and the Caribbean. AIDS research and human retroviruses. 2015;31(5):496-503. doi:10.1089/aid.2014.0241.

53. Orsini N, Li R, Wolk A, Khudyakov P, Spiegelman D. Meta-analysis for linear and nonlinear dose-response relations: examples, an evaluation of approximations, and software. Am J Epidemiol. 2012;175(1):66-73. doi:10.1093/aje/kwr265.

54. Royston P, Altman DG. Approximating statistical functions by using fractional polynomial regression. *J Royal Stat Soc* 1997;46(3):411-422.

55. Jones BL, Nagin DS. A Stata plugin for estimating group-based trajectory models. *Sociological Methods & Research* 2012;42(4):608-613.

56. Huys R, Jirsa VK. Nonlinear dynamics in human behavior. Springer; 2010.

57. Crutzen R, Giabbanelli P. Using Classifiers to Identify Binge Drinkers Based on Drinking Motives. *SUBST USE MISUSE* 2013;49(1-2):110-5. doi:10.3109/10826084.2013.824467.

58. Chen SY, Feng Z, Yi X. A general introduction to adjustment for multiple comparisons. J Thorac Dis. 2017;9(6):1725-9. doi:10.21037/jtd.2017.05.34.

59. Hsu J. Multiple comparisons: theory and methods. Chapman and Hall/CRC; 1996.

60. Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano RP et al. Structure of dietary measurement error: results of the OPEN biomarker study. Am J Epidemiol. 2003;158(1):14-21; discussion 2-6. doi:10.1093/aje/kwg091.

61. Schatzkin A, Kipnis V, Carroll RJ, Midthune D, Subar AF, Bingham S et al. A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based Observing Protein and Energy Nutrition (OPEN) study. International journal of epidemiology. 2003;32(6):1054-62. doi:https://doi.org/10.1093/ije/dyg264.

62. Hedrick VE, Dietrich AM, Estabrooks PA, Savla J, Serrano E, Davy BM. Dietary biomarkers: advances, limitations and future directions. Nutrition journal. 2012;11:109. doi:10.1186/1475-2891-11-109.

63. Dragsted LO, Gao Q, Scalbert A, Vergeres G, Kolehmainen M, Manach C et al. Validation of biomarkers of food intake-critical assessment of candidate biomarkers. Genes Nutr. 2018;13:14. doi:10.1186/s12263-018-0603-9.

64. Crews H, Olivier IV L, Wilson L. Urinary biomarkers for assessing dietary exposure to caffeine. Food additives & contaminants. 2001;18(12):1075-87.

65. Thompson HJ, Heimendinger J, Diker A, O'Neill C, Haegele A, Meinecke B et al. Dietary botanical diversity affects the reduction of oxidative biomarkers in women due to high vegetable and fruit intake. The Journal of nutrition. 2006;136(8):2207-12.

66. Woodside JV, Draper J, Lloyd A, McKinley MC. Use of biomarkers to assess fruit and vegetable intake. The Proceedings of the Nutrition Society. 2017;76(3):308-15. doi:10.1017/s0029665117000325.

67. Ross AB. Present status and perspectives on the use of alkylresorcinols as biomarkers of wholegrain wheat and rye intake. Journal of nutrition and metabolism. 2012;2012.

68. Andersson A, Marklund M, Diana M, Landberg R. Plasma alkylresorcinol concentrations correlate with whole grain wheat and rye intake and show moderate reproducibility over a 2- to 3-month period in free-living Swedish adults. The Journal of nutrition. 2011;141(9):1712-8.

69. Davis CD, Hord NG. Nutritional "omics" technologies for elucidating the role(s) of bioactive food components in colon cancer prevention. J Nutr. 2005;135(11):2694-7. doi:10.1093/jn/135.11.2694.

70. Mead MN. Nutrigenomics: the genome--food interface. Environmental health perspectives. 2007;115(12):A582-9. doi:10.1289/ehp.115-a582.

71. Nicastro HL, Trujillo EB, Milner JA. Nutrigenomics and cancer prevention. Current nutrition reports. 2012;1(1):37-43. doi:10.1007/s13668-011-0007-6.

72. Yang A, Palmer AA, de Wit H. Genetics of caffeine consumption and responses to caffeine. Psychopharmacology (Berl). 2010;211(3):245-57. doi:10.1007/s00213-010-1900-1.

73. Lindon J, Nicholson J. Systems biology metabonomics. Nature. 2008;455:1054-6.

74. McGee EE, Kiblawi R, Playdon MC, Eliassen AH. Nutritional metabolomics in cancer epidemiology: current trends, challenges, and future directions. Current nutrition reports. 2019;8(3):187-201. doi:https://doi.org/10.1007/s13668-019-00279-z.

75. Playdon MC, Ziegler RG, Sampson JN, Stolzenberg-Solomon R, Thompson HJ, Irwin ML et al. Nutritional metabolomics and breast cancer risk in a prospective study. The American journal of clinical nutrition. 2017;106(2):637-49. doi:10.3945/ajcn.116.150912.

76. Chadeau-Hyam M, Athersuch TJ, Keun HC, De Iorio M, Ebbels TM, Jenab M et al. Meeting-in-the-middle using metabolic profiling - a strategy for the identification of intermediate biomarkers in cohort studies. Biomarkers. 2011;16(1):83-8. doi:10.3109/1354750x.2010.533285.

77. Troche JR, Mayne ST, Freedman ND, Shebl FM, Guertin KA, Cross AJ et al. Alcohol Consumption-Related Metabolites in Relation to Colorectal Cancer and Adenoma: Two Case-Control Studies Using Serum Biomarkers. PloS one. 2016;11(3):e0150962. doi:10.1371/journal.pone.0150962.

78. Guertin KA, Loftfield E, Boca SM, Sampson JN, Moore SC, Xiao Q et al. Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. The American journal of clinical nutrition. 2015;101(5):1000-11. doi:10.3945/ajcn.114.096099.

79. Romagnolo DF, Milner JA. Opportunities and challenges for nutritional proteomics in cancer prevention. J Nutr. 2012;142(7):1360s-9s. doi:10.3945/jn.111.151803.

80. Wang J, Li D, Dangott LJ, Wu G. Proteomics and its role in nutrition research. J Nutr. 2006;136(7):1759-62. doi:10.1093/jn/136.7.1759.

81. Fuchs D, Winkelmann I, Johnson IT, Mariman E, Wenzel U, Daniel H. Proteomics in nutrition research: principles, technologies and applications. The British journal of nutrition. 2005;94(3):302-14. doi:10.1079/bjn20051458.

82. Daniel H. Genomics and proteomics: importance for the future of nutrition research. The British journal of nutrition. 2002;87 Suppl 2:S305-11. doi:10.1079/bjnbjn/2002554.

83. Kussmann M, Affolter M, Fay LB. Proteomics in nutrition and health. Combinatorial chemistry & high throughput screening. 2005;8(8):679-96. doi:10.2174/138620705774962526.

84. Kim H. New nutrition, proteomics, and how both can enhance studies in cancer prevention and therapy. J Nutr. 2005;135(11):2715-8. doi:10.1093/jn/135.11.2715.

85. Go VL, Butrum RR, Wong DA. Diet, nutrition, and cancer prevention: the postgenomic era. J Nutr. 2003;133(11 Suppl 1):3830s-6s. doi:10.1093/jn/133.11.3830S.

86. Ayoub HM, McDonald MR, Sullivan JA, Tsao R, Meckling KA. Proteomic profiles of adipose and liver tissues from an animal model of metabolic syndrome fed purple vegetables. Nutrients. 2018;10(4):456. doi:https://doi.org/10.3390/nu10040456.

87. Brauer HA, Libby TE, Mitchell BL, Li L, Chen C, Randolph TW et al. Cruciferous vegetable supplementation in a controlled diet study alters the serum peptidome in a GSTM1-genotype dependent manner. Nutrition journal. 2011;10:11. doi:10.1186/1475-2891-10-11.

88. Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. Oxford University Press; 2002.

Additional Information

Disclaimer

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

Ethics approval and consent to participate

The *Bladder Cancer Prognoses Programme* (BCPP) study involved humans and ethical approval was obtained, ethics reference: 06/MRE04/65. The protocol had been registered in clinicaltrials.gov, registration number: NCT00553215. All participants provided written informed consent before they were included in the study.

Each participating study in the *BLadder Cancer Epidemiology and Nutritional Determinants* (BLEND) consortium has been approved by the local ethic committee. Informed consent was obtained from all individual participants included in each participating study.

Conflict of Interest

The authors declare that they have no conflict of interest.

Funding

The BCPP project is partly funded by Cancer Research UK.

The BLEND project is partly funded by the World Cancer Research Fund International (WCRF 2012/590) and European Commission (FP7-PEOPLE-618308).

Hessen Case–control study on bladder cancer was supported by the Bundesanstalt für Arbeitsschutz (No. F 1287). The Kaohsiung study was supported by grant NSC 85-2332-B-037-066 from the National Scientific Council of the Republic of China. The Stockholm Case-control study was supported by grant from the Swedish National Cancer Society and from the Swedish Work Environment Fund. The Roswell Park Memorial Institute Case-control study on bladder cancer was supported by Public Health Service Grants CA11535 and CA16056 from the National Cancer Institute. The New England bladder cancer study was funded in part by

grant numbers 5 P42 ES007373 from the National Institute of Environmental Health Sciences, NIH and CA57494 from the National Cancer Institute, NIH. The Italian Case-control study on bladder cancer was conducted within the framework of the CNR (Italian National Research Council) Applied Project “Clinical Application of Oncological Research” (contracts 94.01321.PF39 and 94.01119.PF39), and with the contributions of the Italian Association for Cancer Research, the Italian League against Tumours, Milan, and Mrs. Angela Marchegiano Borgomainerio. The Brescia bladder cancer study was partly supported by the International Agency for Research on Cancer. The French INSERM study was supported by a grant from the Direction Générale de la Santé, Ministère des Affaires Sociales, France. The Molecular Epidemiology of Bladder Cancer and Prostate Cancer was supported in part by grants ES06718 (to Z.-F.Z.), U01 CA96116 (to A.B.), and CA09142 from the NIH National Institute of Environmental Health Sciences, the National Cancer Institute, the Department of Health and Human Services, and by the Ann Fitzpatrick Alper Program in Environmental Genomics at the Jonsson Comprehensive Cancer Center, UCLA. The Women’s Lifestyle and Health Study was funded by a grant from the Swedish Research Council (Grant number 521-2011-2955). The Netherlands Cohort Study on diet and cancer was supported by the Dutch Cancer Society. The RERF atomic bomb survivors Study was supported by The Radiation Effects Research Foundation (RERF), Hiroshima and Nagasaki, Japan, a public interest foundation funded by the Japanese Ministry of Health, Labour and Welfare (MHLW) and the US Department of Energy (DOE). The research was also funded in part through DOE award DE-HS0000031 to the National Academy of Sciences. This publication was supported by RERF Research Protocol RP-A5-12. The VITamins and Lifestyle Study (VITAL) was supported by a grant (R01CA74846) from the National Cancer Institute.

The Netherlands Cohort Study on diet and cancer was supported by the Dutch Cancer Society. The RERF atomic bomb survivors Study was supported by The Radiation Effects Research Foundation (RERF), Hiroshima and Nagasaki, Japan, a public interest foundation funded by the Japanese Ministry of Health, Labour and Welfare (MHLW) and the US Department of Energy (DOE). The research was also funded in part through DOE award DE-HS0000031 to the National Academy of Sciences. This publication was supported by RERF Research Protocol RP-A5-12. The VITamins and Lifestyle Study (VITAL) was supported by a grant (R01CA74846) from the National Cancer Institute. The *European Prospective Investigation into Cancer and*

Nutrition (EPIC) was carried out with financial support of the 'Europe Against Cancer' Programme of the European Commission (SANCO); Ligue contre le Cancer (France); Société 3 M (France); Mutuelle Générale de l'Éducation Nationale; Institut National de la Santé et de la Recherche Médicale (INSERM); Institute Gustave Roussy; German Cancer Aid; German Cancer Research Centre; German Federal Ministry of Education and Research; Danish Cancer Society; Health Research Fund (FIS) of the Spanish Ministry of Health; the Spanish Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra; Cancer Research UK; Medical Research Council, UK; Stroke Association, UK; British Heart Foundation; Department of Health, UK; Food Standards Agency, UK; Wellcome Trust, UK; Greek Ministry of Health; Greek Ministry of Education; Italian Association for Research on Cancer; Italian National Research Council; Dutch Ministry of Public Health, Welfare and Sports; Dutch Prevention Funds; LK Research Funds; Dutch ZON (Zorg Onderzoek Nederland); World Cancer Research Fund; Swedish Cancer Society; Swedish Scientific Council; Regional Government of Skane, Sweden; Norwegian Cancer Society; Norwegian Research Council. Partial support for the publication of this supplement was provided by the Centre de Recherche et d'Information Nutritionnelles (CERIN). Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by Cancer Council Victoria [http://www.cancervic.org.au/] and VicHealth [https://www.vichealth.vic.gov.au/]. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database.

Author Contributions

The candidate has performed all the statistical analyses, drafted all manuscripts and engaged in data management activities. Co-authors on the different manuscripts have contributed to this work through supervision, data collection, data management, and providing comments on the drafted manuscripts, giving advice on the statistical analysis or having previously obtained funding needed to perform the study.

Samenvatting

Het doel van deze thesis was inzicht verkrijgen in de *health-related quality of life* (HRQoL) van blaaskankerpatiënten op het moment van diagnose en het vergroten van ons begrip van de invloed van voedselconsumptie op blaaskankerrisico door gebruik te maken van zowel traditionele epidemiologische methoden als nieuwe alternatieven (*i.e.* machine learning). Om dit te bereiken, hebben we de volgende zaken onderzocht: a) de HRQoL van blaaskankerpatiënten in het Verenigd Koninkrijk (VK) rond het moment van diagnose (**hoofdstuk 2**) [gebaseerd op de *Bladder Cancer Patients Prognosis* (BCPP) multi-centre cohort studie]; b) de invloed van koffie (**hoofdstuk 3-4**), groenten (**hoofdstuk 5**), granen en voedingsvezels (**hoofdstuk 6**), en potentiële dieetpatronen (**hoofdstuk 7**), op blaaskankerrisico [gebaseerd op *BLadder cancer Epidemiology and Nutritional Determinants* (BLEND) internationaal consortium].

In **Hoofdstuk 2** toonden we aan dat hoge HRQoL-scores bij patiënten op het moment van of rond de diagnose vergelijkbaar zijn met die van de algemene bevolking van Duitsland en mogelijk het VK. Over het algemeen hadden patiënten met niet-spierinvasieve blaaskanker (NMIBC) een ietwat hogere HRQoL in vergelijking met patiënten met spierinvasieve blaaskanker, wat suggereert dat de progressie van blaaskanker de HRQoL wellicht doet afnemen. Echter, alleen het domein "*Role Functioning*" was klinisch significant hoger onder NMIBC-patiënten. In **hoofdstuk 2** werd tevens onderzocht welke basiskenmerken ten tijde van de diagnose invloed uitoefenen op de HRQoL van blaaskankerpatiënten. Factoren die de HRQoL negatief beïnvloedden waren geslacht (vrouwelijk), hogere leeftijd, gevorderd stadium en een hogere graad van comorbiditeit.

Aangezien blaaskanker kan worden beschouwd als een chronische ziekte, zouden de resultaten van **hoofdstuk 2** van toepassing kunnen zijn bij het bepalen van de waarde van een behandeling en zouden deze verder kunnen dienen als basiswaardes voor toekomstige interventiestudies. Kennis over de HRQoL ten tijde van diagnose stelt men ertoe in staat om veranderingen in de HRQoL als gevolg van de ziekte en behandelingen te meten. Deze kunnen worden gebruikt om het volledige effect van de ziekte en de behandeling in te schatten en om onderscheid te maken tussen behandelingen die in andere opzichten hetzelfde zijn. Dit verhoogt het bewustzijn onder zorgverleners zodat de behoeftes van de patiënt beter kunnen worden ingeschat, wat de klinische resultaten zal verbeteren. Daarnaast vergroot dit het

bewustzijn van blaaskanker, wat kan leiden tot meer gelegenheden voor het verwerven van onderzoeksfinanciering.

Hoofdstuk 2 beschreef de HRQoL van blaaskankerpatiënten ten tijde van diagnose. De resterende hoofdstukken van deze thesis richtten zich op de voedingsgerelateerde risicofactoren voor blaaskanker. In **hoofdstuk 3** en **hoofdstuk 4** werd de associatie tussen koffieconsumptie en blaaskankerrisico onderzocht door middel van het verzamelen van data van respectievelijk case-control- en cohortonderzoeken. De resultaten van deze hoofdstukken zijn tegenstrijdig; **hoofdstuk 3** toonde namelijk een verhoogd risico op blaaskanker bij hogere koffieconsumptie onder nooit-rokers maar niet onder rokers, terwijl de resultaten van hoofdstuk 4 een verhoogd blaaskankerrisico aantoonde bij hogere koffieconsumptie onder mannelijke rokers.

Het wordt al lang erkend dat er een sterke positieve associatie bestaat tussen roken en koffieconsumptie, zowel observationeel als genetisch. Daarnaast wordt roken gezien als de meest belangrijke confounder bij onderzoek naar de associatie tussen koffieconsumptie en ziekterisico. Hierdoor kan het onafhankelijke effect van koffieconsumptie op ziekterisico alleen worden gevonden als roken volledig als confounder wordt geëlimineerd. Resultaten van **hoofdstuk 3** en **hoofdstuk 4** zijn mogelijk beïnvloed door residual confounding van de rookhistorie. De case-controlestudies in **hoofdstuk 3** hadden een retrospectieve dataverzameling over de voorgaande 1-2 jaren. Hierdoor missen ze essentiële informatie over rookgedrag op jongere leeftijd, wat mogelijk gevolgen heeft voor blaaskankerrisico. In tegenstelling hierop heeft **hoofdstuk 4** alleen informatie over de rookstatus aan het begin van de studie, waardoor hier mogelijk essentiële informatie mist over later rookgedrag. Daarnaast hebben sommige personen, met name vrouwen en jong-volwassenen, de neiging om rookgedrag te onderrapporteren, aangezien roken als sociaal onwenselijk wordt beschouwd.

Hoofdstuk 5 geeft bewijs voor een beschermend effect van totale groenteconsumptie, koolhydraatarme groenteconsumptie en boerenkoolconsumptie op blaaskankerrisico onder vrouwen. Hoewel het nutritionele belang van groenten al lang wordt erkend onder voedingsdeskundigen en de medische gemeenschap en er vele voedingsrichtlijnen bestaan die een hoge inname van groenten als belangrijk component bevatten, werd in het gehele

cohort geen beschermend effect gevonden op blaaskankerrisico van de totale inname of (sub)types van groenten, met als uitzondering boerenkool.

Graanconsumptie speelt een belangrijke rol voor voedingspatronen over de hele wereld en de gezondheidseffecten van graanconsumptie, met name volkoren granen, ziet veel onderzoeksinteresse. Bewijs voor het effect op blaaskankerrisico mist echter momenteel nog. Daarom hadden we in **Hoofdstuk** 6 het doel om het effect van volkoren en geraffineerde graanconsumptie op blaaskanker te onderzoeken. Hier vonden we dat een hogere consumptie van volkoren granen het risico op blaaskanker verlaagde, terwijl er geen associatie werd gevonden voor de totale consumptie van geraffineerde granen. Aangezien voedingsvezels een belangrijk verschil vormen tussen volkoren en geraffineerde vezels, werd de relatie tussen voedingsvezels en blaaskankerrisico ook onderzocht. Deze werd omgekeerd geassocieerd bevonden. Deze resultaten ondersteunen huidige adviezen om de inname van volkoren graanproducten en voedingsvezels te verhogen voor de verbetering van de volksgezondheid.

Aangezien voedsel en dranken samen worden genuttigd, werd in **Hoofdstuk 7** een meer holistische aanpak gebruikt. Met behulp van machine learning werden voedselgroepen geïdentificeerd en geprioriteerd die gerelateerd zijn aan blaaskankerrisico. De resultaten van dit hoofdstuk onthulden vijf voedselgroepen gerelateerd aan blaaskankerrisico (d.w.z. op volgorde van belang: granen en graanproducten; groenten en groenteproducten; dranken (niet-melk); vetten; oliën en olieproducten, vlees en vleesproducten).

Valorisation Addendum

Bladder cancer, the 4th most common cancer in men and 17th most common in women [1, 2], represents a major health problem. It is therefore, of utmost importance to establish risk factors associated with the development of this disease.

The USA National Cancer Institute has stated that one-third of all bladder cancer deaths could have been prevented by dietary modifications. A 30 % reduction of bladder cancer incidence in the USA would equal a reduction of 1.2 billion US dollars of total medical care expenditures per year. This thesis provides clues to the dietary determinants of bladder cancer, thereby providing data about the diet-bladder cancer relationship, that can directly be transformed Public Health Nutrition into the practise of prevention. Even though the effect sizes of the food items investigated in this thesis are small, they might have a major impact on human health and related health care costs.

A major challenge remains in translating nutritional findings into daily practice. Although in recent years dietary guidance has become increasingly science based, there seems to be an ever-widening gap between the scientific evidence and consumer behaviour. Consumers are overwhelmed with dietary recommendations they receive from television, popular press, or by browsing internet, making it hard to make the accurate nutritional decisions. It is the task for scientific experts and government to fill this gap and provide individuals with accurate evidence-based nutritional recommendations. For this, close collaboration and knowledge exchange between scientists, health professionals like dietitians and nutritionists, health care organization and governments are needed.

In order to meet this need, results of the current thesis have been not only been published in scientific journals, but have also been presented to the academic community during well-known scientific conferences [*i.e.* yearly Dutch Epidemiological Society (WEON)] and disseminated to the public by an Elsevier (with the endorsement of the WCRF) published book entitled “Diet and Fighting Bladder Cancer”. This book specifically aims to help (layman) individuals to make accurate nutritional decisions to fight bladder cancer by translating all current scientific findings (including ours) into practical and example recipes. In addition, a lay version of our results will be published on wikipedia, e-how/health and the project’s own website (www.dietandbladdercancer.org).

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA: a cancer journal for clinicians. 2020;70(1):7-30. doi:https://doi.org/10.3322/caac.21590.

2. Estimated age-standardized incidence rates (World). Available at: https://gcoiarcfr/today/online-analysis-multi-bars?v=2018&mode=cancer&mode_population=countries&population=900&populations=900&key=asr&sex=2&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&nb_items=50&group_cancer=1&include_nmsc=1&include_nmsc_other=1&type_multiple=%257B%2522inc%2522%253Atrue%252C%2522mort%2522%253Afalse%252C%2522prev%2522%253Afalse%257D&orientation=horizontal&type_sort=0&type_nb_items=%257B%2522top%2522%253Atrue%252C%2522bottom%2522%253Afalse%257D&population_group_globocan_id=. Accessed May 19th, 2020.