


# Investigating Sex Differences in Cancer Incidence

**Maira Khan**

Lady Margaret Hall College

University of Oxford

A thesis submitted for the degree of

*Doctor of Philosophy*

Michaelmas 2025-26

# Abstract

**Background** Sex differences in cancer incidence are well documented, with men generally at higher risk for most cancers shared by both sexes. Whether these disparities reflect differences in risk factor distributions or underlying biological mechanisms remains unclear. Much of the existing evidence is based on registry statistics, with relatively few cohort-based analyses using individual-level data. This thesis addresses these gaps by investigating the contributions of established and emerging risk factors to sex differences in cancer incidence.

**Methods** Prospective analyses were conducted in the UK Biobank cohort using Cox proportional hazards models to examine associations between sex and risk of 15 cancers (and 13 subtypes). Statistical models were stratified by age, deprivation index, and region, with multivariable adjustment for ethnicity, educational attainment, height, BMI, smoking, alcohol, and site-specific risk factors. Subsequent analyses examined adiposity traits beyond BMI in relation to oesophageal adenocarcinoma (OAC), rectal, and kidney cancers to explore potential contributors for the higher risk in men. The greater risk of lung adenocarcinoma (LUAD) in never smokers in women was also investigated. Contributions of risk factors that differed in prevalence by sex (household fuel use and asthma) and sex-specific factors (menopausal status) were evaluated.

**Results** Over 10.5 (SD 2.2) years of follow-up, 32,315 incident cancers across 15 anatomical sites (58.1% in women) were identified among 470,771 individuals (53.8% women). Some sex differences in cancer risk attenuated to the null in the multivariable-adjusted models, but men remained at higher risk than women for eight cancers; the largest differences were for OAC (HR 5.45, 95% CI 4.18–7.12), gastric cardia (HR 3.65, 2.48–5.38), bladder (HR 3.47, 2.85–4.24), liver (HR 1.91,1.48-2.47), rectum (HR 1.70, 1.47–1.96), and kidney cancer (HR 1.77, 1.51–2.09). Adjustment for waist-to-hip ratio attenuated the sex difference in kidney cancer (HR 1.36, 1.13–1.65). Conversely, women had higher incidence for four sites, including LUAD (HR 0.72, 0.62–0.84). The higher LUAD risk in women persisted among never smokers, even after exploration of contributory risk factors.

**Summary** Using data from a large UK cohort study, sex differences in cancer incidence were identified that are only partially explained by cancer risk factors. Four cancer sites—oesophageal adenocarcinoma, bladder, and rectum (higher incidence in men), and lung adenocarcinoma (higher incidence in women)—had marked sex differences that remained upon evaluation of contributions from risk factors and therefore merit further investigation. This thesis demonstrates that systematic cohort-wide investigation of sex differences in cancer incidence can enhance understanding of cancer aetiology.

# Acknowledgments

I express my sincerest gratitude to my supervisors, Professor Ruth Travis, Professor Tim Key, and Dr Joshua Atkins, whose unwavering enthusiasm, support, and generosity with their time have been pivotal to this research. Every supervision meeting led to interesting questions and new ideas to consider when thinking about sex differences in cancer epidemiology. I thank them for taking me on as their student; it has been a privilege and the highlight of my time at Oxford to have learnt under them.

I thank and shall always be indebted to my Mum, Dr Sharmin Yaqin, whose career dedicated to oncology and public health in India has inspired my education. I have been incredibly fortunate to witness her commitment to caring for cancer patients, to working hard, and to loving her career. I thank my siblings and my best friends, Arham and Daanya, who have been endlessly loving and supportive. I thank my grandparents, Professor Anwarul Yaqin and Ms Badarjahan Yaqin, for their kindness and love for me and my family.

I am grateful to be a part of the Cancer Epidemiology Unit, and to have worked with and learnt from the many people in this team. Thank you to Kirstin Pirie, Dr Keren Papier, Professor Tammy Tong, Professor Karl Smith-Byrne, Dr Kezia Gaitskell, Professor Isobel Barnes, and Professor Gillian Reeves. I also thank Dixa Thakrar, Tonima Trisa, Zoe Grenville, Yang Tao, and Xuan Li for sharing their DPhil journeys with me. I am also grateful to have been part of the Nuffield Department of Population Health, and to have had the opportunity to serve on the DPhil student committee and have worked with Marie-Alice Frappat, Professor Mark Sheehan, and Professor David Preiss.

I also acknowledge my mentors and supervisors from Cambridge University—Professor Antonis Antoniou, Dr Nasim Mavaddat, Dr Stephen Burgess, and Professor John R. B. Perry—and from Royal Holloway, Professor Vincent Jansen. Their mentorship and advice were instrumental in shaping my academic interests and career in public health.

I thank my friends—Diego, Anju, Billy, Ash, Louise, Mathias, Bowen, Charlotte, and Yangyang—who have persistently shown unwavering care and support for me. It has been a tremendous joy to have studied and lived at Oxford these last four years; to have made so many wonderful friends here and to be a member of the Oxford University Amateur Boxing Club and Lady Margaret Hall College—something I will always cherish.

# Data Availability and Ethics Declaration

The UK Biobank resource used for the doctoral work is available for access to approved health researchers who may use the resource for health-related research in the public interest. The study was approved by the North West Multi-Centre Research Ethics Committee (UK) (reference number 06/MRE08/65).

At recruitment, all participants gave informed consent to participate and for their health to be followed-up through linkage to their health-related records. All methods were undertaken in accordance with the relevant guidelines and regulations. More information on the UK Biobank Access Management System can be found through https://www.ukbiobank.ac.uk/register-apply/.

I express my sincerest gratitude to the participants who contributed to the UK Biobank resource, and the researchers and staff who established and have maintained it. Without their time and effort, this research would have not been possible to conduct.

# DPhil Funding and Output

My DPhil journey has been supported by the Nuffield Department of Population Health Studentship, and I thank the Nuffield Department of Population Health for awarding me the scholarship to undertake this doctoral work.

Chapter 4 of this thesis has been adapted and published in the British Journal of Cancer (https://doi.org/10.1038/s41416-025-03028-y) (2025).

# Abbreviations

| | |
|---|---|
| ASAT | Abdominal subcutaneous adipose tissue |
| ASIR | Age-standardised incidence rate (per 100,000 person-years) |
| ATV | Adipose tissue volume |
| BFP | Body fat percentage |
| BMI | Body mass index |
| CI | Confidence interval |
| CKB | China Kadoorie Biobank Study |
| CRC | Colorectal cancer |
| EPIC | European Prospective Investigation into Cancer and Nutrition |
| FDA, US | US Food and Drug Administration |
| GWAS | Genome-wide association study |
| HC | Hip circumference |
| Hep B or C | Hepatitis B or C |
| HIV | Human immunodeficiency virus |
| HPV | Human papillomavirus |
| HR | Hazard ratio |
| HL | Hodgkin lymphoma |
| IARC | International Agency for Research on Cancer |
| ICD | International Classification of Diseases |
| LCINS | Lung cancer in never smokers |
| LSCC | Lung squamous cell carcinoma |
| LUAD | Lung adenocarcinoma |
| IVW | Inverse variance weighted |
| MET | Metabolic equivalent |
| MF IRR | Male-to-female incidence rate ratio |
| MR | Mendelian randomization |
| MWS | Million Women Study |
| NCI, US | US National Cancer Institute |
| NHL | Non-Hodgkin lymphoma |
| NHS | National Health Service |
| NIH, US | US National Institutes of Health |
| NIH-AARP | NIH-American Association of Retired Persons |

| | |
|---|---|
| NIHR, UK | UK National Institute for Health and Care Research |
| NSCLC | Non-small cell lung carcinoma |
| OAC | Oesophageal adenocarcinoma |
| ONS | Office for National Statistics |
| OPCS-4 | Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures (version 4) |
| OR | Odds ratio |
| ORWH | Office of Research on Women's Health |
| OSCC | Oesophageal squamous cell carcinoma |
| RCC | Renal cell carcinoma |
| RR | Relative risk |
| SABV | Sex as biological variable |
| SAGER | Sex and Gender Equity in Research |
| SCC | Squamous cell carcinoma |
| SCLC | Small cell lung carcinoma |
| SEER, US | US Surveillance, Epidemiology, and End Results Program |
| TFP | Trunk fat percentage |
| UK | United Kingdom |
| UKB | UK Biobank |
| US | United States |
| UV | Ultraviolet |
| VAT | Visceral adipose tissue |
| WC | Waist circumference |
| WHI | Women's Health Initiative |
| WHR | Waist-to-hip ratio |
| WHtR | Waist-to-height ratio |

# Table of Contents

**Chapter 1 Introduction 12**

*1.1 Overview 12*

*1.2 Research Rationale and Significance 16*

*1.3 Epidemiological Concepts and Definitions 21*

*1.4 Central Aim and Analytical Approach 23*

*1.5 Thesis Objectives and Structure 25*

**Chapter 2 A Review of the Sex Differences in Cancer Incidence 29**

*2.1 Trends for Sex Differences in Cancer Incidence 29*

2.1.1 Temporal Trends 29

2.1.2 Age-related Trends 31

*2.2 Sex Differences Across Extrinsic Risk Factors 35*

2.2.1 Sociodemographic Characteristics 35

2.2.2 Smoking 38

2.2.3 Alcohol 39

2.2.4 Infections 41

2.2.5 Occupations 42

*2.3 Sex Differences Across Intrinsic Risk Factors 44*

2.3.1 Height 44

2.3.2 Body Mass Index 45

2.3.3 Genetic Factors 48

2.3.4 Sex Hormones 51

*2.4 Sex Differences in Incidence by Cancer Site 54*

2.4.1 Oral cavity (C00-C14) 55

2.4.2 Oesophagus [C15] 56

2.4.3 Stomach [C16] 58

2.4.4 Colorectum [C18-C20] 60

2.4.5 Anus [C21] 61

2.4.6 Liver [C22] 62

2.4.7 Gallbladder [C23] 63

2.4.8 Pancreas [C25] 64

2.4.9 Lung [C34] 65

2.4.10 Malignant melanoma [C43] 67

2.4.11 Kidney [C64-C65] 68

2.4.12 Bladder [C67] 69

2.4.13 Thyroid [C73] 70

2.4.14 Lymphatic and Haematopoietic Cancers [C81-C96] 71

*2.5 Summary 75*

**Chapter 3 UK Biobank Cohort: Design, Variables, and Principal Statistical Approach 77**

*3.1 UK Biobank Cohort Design and Baseline Assessment 77*

*3.2 Analytic Cohort Characteristics 78*

*3.3 Exclusion Criteria* *84*

*3.4 Exposure: Sex* *85*

*3.5 Ascertainment of Cancer Incidence* *85*

*3.6 Cancer Endpoints and Case Counts* *85*

*3.7 Statistical Power Calculations* *88*

*3.8 Principal Cox Regression Modelling Approach* *90*

*3.9 Variable Coding for Baseline Multivariable-adjusted Analyses* *92*

3.9.1 Age at Recruitment 92

3.9.2 Region 92

3.9.3 Townsend Deprivation Index 93

3.9.4 Ethnicity 93

3.9.5 Qualifications 94

3.9.6 Height 94

3.9.7 Body mass index (BMI) 94

3.9.8 Smoking 95

3.9.9 Alcohol 96

*3.10 Summary* *97*

**Chapter 4 Sex Differences in Cancer Incidence: Prospective Analyses in the UK Biobank 99**

*4.1 Introduction* *99*

*4.2 Methods* *101*

4.2.1 Cancer Site-specific Multivariable-adjusted Cox Models 101

4.2.2 Sensitivity Analyses 103

*4.3 Results* *105*

4.3.1 UK Biobank Participants 105

4.3.2 Associations Between Sex and Cancer Endpoints 105

4.3.3 Sensitivity Analyses 106

*4.4 Discussion* *120*

4.4.1 Oesophageal adenocarcinoma 121

4.4.2 Bladder 121

4.4.3 Kidney 123

4.4.4 Rectum 123

4.4.5 Lung adenocarcinoma 124

4.4.6 Thyroid 125

4.4.7 Strengths, Limitations and Conclusions 125

**Chapter 5 Exploring Contributions of Adiposity Traits to Sex Differences in Oesophageal adenocarcinoma, Kidney, and Rectal Cancer Risk 128**

*5.1 Introduction* *128*

*5.2 Methods* *131*

5.2.1 Cohort Exclusions and Adiposity Trait Variables 131

5.2.2 Cross-sectional Analyses of Adiposity Measures 132

5.2.3 Statistical Analyses 133

*5.3 Results* *135*

5.3.1 UK Biobank Participants 135
5.3.2 Cross-sectional Associations Between Adiposity Traits 136
5.3.3 Sex-stratified Associations for Adiposity Traits and Cancers 137
5.3.4 Sex Differences in Cancer Risk by Adiposity Traits 139

*5.4 Discussion* *153*

**Chapter 6 Lung adenocarcinoma in Never Smokers: Investigating Hypotheses for Increased Risk in Women** **156**

*6.1 Introduction* *156*
6.1.1 Epidemiology 157
6.1.2 Sex Differences in Incidence Beyond Smoking 158
6.1.3 Established LUAD Risk Factors in Never Smokers 160
6.1.4 Relevance of Risk Factors towards Sex Differences 161
6.1.5 Research Objective 164

*6.2 Methods* *167*

*6.3 Results* *168*

*6.4 Discussion* *173*

**Chapter 7 Discussion and Conclusions** **174**

*7.1 Key Findings and Research Directions* *174*

*7.2 Strengths and Research Contributions* *177*

*7.3 Limitations* *178*

*7.4 Further Research Considerations* *180*

*7.5 Public Health Implications (NOTES!)* *186*

*7.6 Conclusions* *188*

**Appendices 192**

**References 220**

# List of Figures

**Figure 1.1 Conceptual representation of the contributions of extrinsic and intrinsic risk factors across stages of carcinogenesis. ..... 27**

**Figure 1.2 Classification of extrinsic and intrinsic risk factors in epidemiology hypothesized to contribute to sex differences in cancer incidence. ..... 28**

**Figure 2.1 Age standardised incidence rates per 100,000 from 1971-2017 for cancers common to both men and women aged 35 and above in the UK (England). ..... 33**

**Figure 2.2 Cancer incidence rates per 100,000 from 2013-2017 for cancers common to both men and women aged 35 and above in the UK (England). ..... 34**

**Figure 2.3 Alcohol consumption levels in the UK for men and women (2012-2022). ..... 41**

**Figure 4.1 Hazard ratios (HR) and 95% confidence intervals (CI) for incident cancer in men compared to women in UK Biobank. ..... 109**

**Figure 5.1 Hazard ratio plots for associations between adiposity traits and oesophageal adenocarcinoma [C15]. ..... 149**

**Figure 5.2 Hazard ratio plots for associations between adiposity traits and kidney cancer [C64-C65]. ..... 150**

**Figure 6.1 Conceptual directed acyclic graph (DAG) illustrating how sex-related differences in the distributions of established risk factors may contribute to observed differences in lung adenocarcinoma (LUAD) incidence among never smokers. ..... 166**

# List of Tables

**Table 2.1 Body mass index (BMI) for adults in the UK (England) by age and sex. ..... 48**

**Table 2.2 Cancer incidence age-standardized rates per 100,000 person-years, for men and women, aged 35 years and above, diagnosed in UK between 2013-2017. ..... 54**

**Table 3.1 Baseline distribution of covariates and details of follow-up in the UK Biobank. ..... 82**

**Table 3.2 Cancer case counts for the 28 selected endpoints in the UK Biobank analytic cohort of 470,771 participants. ..... 87**

**Table 3.3 Statistical power calculations for the 28 selected cancer endpoints in the UK Biobank analytic cohort. ..... 89**

**Table 4.1 Hazard ratios (HRs) and 95% confidence intervals (CIs) comparing risk of cancer in men to women. ..... 110**

**Table 4.2 Cancer specific multivariable-adjusted Cox proportional hazards models comparing risk of cancer incidence in men versus women who identified as never smokers at recruitment in the UK Biobank. ..... 114**

**Table 4.3 Cancer specific multivariable-adjusted Cox proportional hazards models comparing risk of cancer incidence in men versus women who identified as light alcohol drinkers (<20 g/day) in the UK Biobank. ..... 116**

**Table 4.4 Cancer specific multivariable-adjusted Cox proportional hazards models comparing risk of cancer incidence in men versus women had normal BMI at baseline (<25 kg/m2) in the UK Biobank. ..... 118**

**Table 5.1 Sex differences in risk of OAC and cancers of the kidney and rectum for men compared to women (summary of results from Chapter 4; Table 4.4). ..... 131**

**Table 5.2 Distribution of 462,765 UK Biobank participants (53.9% women) by body mass index (BMI) measures at baseline and selected sociodemographic factors. ..... 141**

**Table 5.3 Pearson correlation coefficients between baseline adiposity traits in 213,163 men from the UK Biobank. ..... 143**

**Table 5.4 Pearson correlation coefficients between baseline adiposity traits in 249,602 women from the UK Biobank. ..... 143**

**Table 5.5 Baseline adiposity traits categorised into sex-stratified quartiles in 462,765 UK Biobank participants. ..... 144**

**Table 5.6 Hazard ratios (HRs) and 95% confidence intervals (CIs) for oesophageal adenocarcinoma [ICD-10 C15], in 249,602 women and 213,163 men according to adiposity traits in the UK Biobank. ..... 145**

**Table 5.7 Hazard ratios (HRs) and 95% confidence intervals (CIs) for kidney cancer [ICD-10 C64-C65], in 249,602 women and 213,163 men according to adiposity traits in the UK Biobank. ..... 147**

**Table 5.8 Hazard ratios (95% confidence intervals) comparing men with women for oesophageal adenocarcinoma [ICD-10 C15], adjusted for one adiposity trait at a time. ..... 151**

**Table 5.9 Hazard ratios (95% confidence intervals) comparing men with women for risk of kidney cancer [ICD-10 C64-C65], adjusted for one adiposity trait at a time. ..... 152**

**Table 6.1 Sex differences in risk of lung adenocarcinoma for men compared to women (results from Chapter 4; Table 4.2)..... 157**

**Table 6.2 Characteristics of 257,390 participants of the UK Biobank who self-identified as never smokers at study recruitment..... 170**

**Table 6.3 Household fuel type categorisation in 257,390 never smokers in the UK Biobank..... 172**

**Table 6.4 Hazard ratios (HRs) and 95% confidence intervals (CI) for associations between household fuel type and lung adenocarcinoma [C34] risk, in 149,577 women and 105,223 men from the UK Biobank. ..... 172**

**Table 6.5 Prevalent asthma in 257,390 never smokers in the UK Biobank. ..... 172**

**Table 6.6 Hazard ratios (HRs) and 95% confidence intervals (CI) for associations between prevalent asthma and lung adenocarcinoma [C34] risk, in 150,996 women and 106,394 men from the UK Biobank. ..... 172**

**Table 7.1 Summary of hazard ratios (HR, 95% CI) comparing men to women for cancer sites with markedly[a] higher risk in men. ..... 190**

# Chapter 1 Introduction

## 1.1 Overview

Sex differences in cancer incidence are well documented across populations, with men generally at higher risk than women for cancers that are common to both sexes (International Agency for Research on Cancer, 2025a, National Cancer Institute, 2025, National Disease Registration Service, 2025). In men, the incidence rates are often 1.5- to 3- fold greater than in women, with cancers of the bladder, oesophagus, oral cavity, and liver each having a nearly threefold higher incidence (Cook et al., 2009, Edgren et al., 2012, Jackson et al., 2022). Notably, there are exceptions for certain sites where women exhibit higher rates than men, particularly cancers of the thyroid, gallbladder, and anus (as well as a large difference for breast cancer) (Cook et al., 2009, Edgren et al., 2012, Jackson et al., 2022). There has also been a rise in lung adenocarcinoma in women, with incidence rates surpassing men, especially among never smokers (Jemal et al., 2018, Jemal et al., 2025, Lopiccolo et al., 2024). A study exploring sex differences in cancer incidence using the US Surveillance, Epidemiology, and End Results Program (SEER) data from 1974-2004 reported male-to-female incidence rate ratios across all ages combined were 1.37 (95% CI 1.37-1.38) for all cancer sites, 1.14 (95% CI 1.14-1.14) for all cancer sites excluding sex-specific cancer sites, and 1.77 (95% CI 1.76-1.77) for all cancer sites excluding sex-specific sites and breast cancer (Cook *et al.*, 2009).

Cancer is the second leading cause of death globally and its prevention is one of the most significant challenges in the 21st century (World Health Organization, 2026). Worldwide, an estimated 20 million new cancer cases (10.3 million in men and 9.7 million in women) were diagnosed and 9.7 million cancer deaths (5.4 million in men and 4.3 million in women)

occurred in 2022 (Ferlay *et al.*, 2024). The most common cancers diagnosed globally in 2022 among both sexes combined lung (12.4% of all new cancer cases diagnosed), followed by breast (11.5%), colorectum (9.6%) and prostate (7.3%) and stomach (4.3%). In men, they were lung (15.2% of all new cancer cases diagnosed), followed by prostate (14.2%), colorectum (10.4%) and stomach (6.1%) and liver (5.8%). In women, they were breast (23.8% of all new cancer cases diagnosed), followed by lung (9.4%), colorectum (8.9%), cervix uteri (6.9%) and thyroid (6.4%). By 2050, it is estimated there may be 35.5 million (76.5% higher than 2022) new cancer cases, 19.0 million (84.4% higher) in men and 16.3 million (68.0% more) in women (Ferlay *et al.*, 2024).

In an analysis of the cancers diagnosed globally in 2022, it is estimated approximately 37.8% (7.1 of 18.7 million) of cases were attributable to modifiable risk factors across both sexes globally (Fink *et al.*, 2026). Furthermore, the burden differs substantially by sex, with 4.3 million (45% of 9.6 million) new cancer cases in men attributable to modifiable risk factors compared to 2.7 million (29.7% of 9.2 million) new cancer cases in women (Fink *et al.*, 2026). These disparities are largely attributable to variation in behavioural risk factors that differ across regions. Smoking remains the largest contributor to cancer burden among men globally, whereas smoking and high BMI are key contributors among women in high-income regions (Fink *et al.*, 2026).

Cancer incidence and mortality in the UK are higher than most of the world, estimated to rank 13$^{th}$ for incidence and 65$^{th}$ for mortality out of 185 countries (Cancer Research UK, 2026). The four most common cancers in the UK are female breast, colorectal, lung and prostate, same as the global patterns (Cancer Research UK, 2026). Between 2016 and 2018, age standardised incidence rates for all cancers among adults aged 35–69 in the UK (excluding the two sites with the largest incidence, breast and prostate) were 422.3 per 100,000 for men and 363.8 per 100,000 for women (Shelton et al., 2024). Over the last 25 years these cancer

incidence rates have initially declined and then stabilized in men, whereas a modest annual increase has been observed in women (Shelton et al., 2024). The cancer sites with rising incidence rates in both sexes include cancers of the oral cavity, liver, kidney and melanoma, along with lung cancer in women, all sites which exhibit well documented sex differences (Cook et al., 2009, Edgren et al., 2012, Jackson et al., 2022). While smoking rates are declining in the UK, they remain the top contributor to cancer incidence overall (Brown et al., 2018). Furthermore, the rise in cancer incidence for most sites has been attributed to the rising prevalence of overweight and obesity, which is now the second highest contributor to cancer cases (Brown et al., 2018).

The differences in cancer incidence between sexes have been attributed to differences in exposure to established risk factors, which are largely anthropometric (height and obesity), lifestyle related (smoking, alcohol, diet, and physical activity), molecular (chronic inflammatory and endocrinological risk factors) and genetic (EXITS genes, mosaic loss of the Y chromosome) (Bernardez *et al.*, 2025, Clocchiatti *et al.*, 2016, Dong *et al.*, 2020, Haupt *et al.*, 2021, Jackson *et al.*, 2022, Radkiewicz *et al.*, 2017, Rubin, 2022). These exposure differences are thought to drive a large share of the observed incidence gap; however, a comprehensive understanding of the extent remains to be elucidated. Statistics on sex differences in cancer incidence are usually reported on a national registry level and therefore do not account for hypothesised individual-level cancer risk factors (Cook *et al.*, 2009, Dong *et al.*, 2020, Edgren *et al.*, 2012, Radkiewicz *et al.*, 2017, Scelo *et al.*, 2018, Tosakoon *et al.*, 2024, Zheng *et al.*, 2019). There have also been limited large-scale cross-cancer-site prospective cohort analyses of sex differences in cancer risk. A recent pan-cancer analysis quantified sex disparities for 21 cancer sites in the US National Institutes of Health-American Association of Retired Persons (NIH-AARP) Diet and Health prospective cohort study of over 294,000 participants (41.7% women) (Jackson et al., 2022). Using Cox proportional hazards regression models, the authors assessed the adjusted effect of sex on cancer incidence with

risk factors selected based on their univariate associations with cancer endpoints. They reported that the higher risk in men remained for 20 cancer sites after adjusting for a broad range of risk factors, and only a modest proportion (ranging from 11% to 50%) was explained by those risk factors (Jackson et al., 2022). Their work provides an opportunity to examine the question using an *a priori* approach and leveraging data from a cohort with high statistical power and detailed phenotypic data.

Each cancer with sex differences in its incidence may be attributable to inherent biological differences between males and females, differences in exposure to modifiable risk factors attributable to the cancer, or a combination of both. This thesis investigates sex differences in cancer incidence of 15 cancers (and 13 subtypes) by quantifying the contributions of established risk factors, as they may provide insights into cancer aetiology. This is achieved using data from the UK Biobank, a large prospective cohort study of over 500,000 participants recruited across the United Kingdom (UK), with extensive baseline phenotypic and lifestyle data and longitudinal follow-up through linkage to national health records.

The research focus of this thesis is on sex differences in cancer incidence, examined through biological sex as a variable. Sex is defined as male or female, typically assigned at birth, and determined by genetic, hormonal, and anatomical factors (Office for National Statistics, 2019). Gender, by contrast, is a social construct and relates to how individuals perceive and express their identity, which may not always align with their sex assigned at birth (Office for National Statistics, 2019). Therefore, the findings may not be generalisable to gender-based differences in cancer incidence. Throughout this thesis, the terms “men” and “women” are used to describe participants and are intended to correspond to biological sex as male and female. This terminology reflects a preference that acknowledges subjects as people rather than categories.

## 1.2 Research Rationale and Significance

Studying sex differences in cancer holds relevance across clinical and public health domains. Historically, women were underrepresented in medical research, leading to significant knowledge gaps (Merkatz, 1998, Office of Research on Women's Health, 2024). In 1977, the US Food and Drug Administration (FDA) advised excluding women of childbearing potential from early-phase drug trials to prevent foetal drug exposure (Office of Research on Women's Health, 2024). It led to a large omission of women from clinical research for over a decade and, as a result, many drugs and treatments were developed and dosed based on data using largely men (Office of Research on Women's Health, 2024). In 2001 the US Government Accountability Office withdrew eight out of ten prescription drugs flagged for concern from the markets in the late 1990s as they had greater adverse effects in women than in men (U.S. Government Accountability Office, 2001).

In response to this, the US National Institutes of Health (NIH) Revitalization Act of 1993 mandated the inclusion of women (and additionally minority groups) in all NIH-funded clinical studies (Institute of Medicine (US), 1994). The NIH also established an Office of Research on Women's Health (ORWH) in 1990 to advocate for women's health research (Institute of Medicine (US), 1994, Office of Research on Women's Health, 2024). They commissioned a report to the Institute of Medicine Committee on Understanding the Biology of Sex and Gender Differences titled "Does Sex Matter?" (2001). The committee's purpose was to review and evaluate the current knowledge and scientific evidence related to sex differences in health and illness. They concluded that "yes, sex matters" and being male or female is an important variable that should be considered when designing and analysing studies in biomedical and health research (Institute of Medicine (US), 2001).

This growing recognition led to more global initiatives; in 2016 the Sex and Gender Equity in Research (SAGER) guidelines were introduced to improve the reporting of sex-disaggregated

data in research, and the NIH rolled out a formal policy requiring grant applications to consider sex as a biological variable (SABV) in study designs and analyses (Heidari *et al.*, 2016, Office of Research on Women's Health, 2025). In the UK, the National Institute for Health Research (NIHR) now expects researchers to include diverse populations and analyse outcomes by sex (reinforced further by its 2020 'INCLUDE' guidance and the 2022 UK Women's Health Strategy), and Horizon Europe requires applicants to address sex and gender in research by default (European Commission, 2025, National Institute for Health and Care Research, 2020, National Institute for Health and Care Research, 2025). These efforts have not only increased sex and gender inclusivity in research but also efforts for understanding sex and gender differences in biology and public health.

Alongside these policy implementations, there has been growing interest in using large prospective cohort studies to understand sex differences in epidemiology. Historically, the larger cohort studies were generally single sex, which constrained direct comparisons between men and women. For instance, the British Doctors Study (UK, 1951) and the Physicians' Health Study (US, 1982) exclusively followed up male physicians (Doll and Hill, 1954, Hennekens and Eberlein, 1985, Physicians' Health Study, 1989). The British Doctor's study was initiated as the first large prospective study to investigate the links between tobacco consumption and cause-specific mortality and included both men and women (Doll et al., 1994). However, as the number of women enrolled were low (6,207 women vs. 34,494 men), and tobacco consumption levels were also comparatively lower, they were subsequently excluded from later analyses (Doll et al., 1994). To counterbalance efforts, other studies such as the Nurses' Health Study (US, 1976), Women's Health Initiative (1992), and Million Women Study (UK 1996) were focused on exploring women's health and epidemiology (Colditz and Hankinson, 2005, Million Women Study Collaborative Group, 1999, The Women's Health Initiative, 1998). The sex-specific designs of these studies limited research to be done on men and women comparatively. The establishment of large cohort

studies such as the US NIH-AARP Diet and Health Study (1995), European Prospective Investigation into Cancer (EPIC, 2000s), the UK Biobank (2006), the China Kadoorie Biobank (CKB, 2004), and others have enabled more direct analyses of sex differences in the same study populations (Chen *et al.*, 2005, Riboli, 1997, Schatzkin *et al.*, 2001, Sudlow *et al.*, 2015).

Prior work on cancer epidemiology also provides context for investigating sex differences. Arguably one of the earliest studies of sex differences in cancer incidence and the aetiology of cancer remains the work on smoking as a cause of lung cancer (Peto, 2023). The research on smoking causing over a third of all cancer deaths began in the 1950s from noting that an increasing proportion of certified causes of death were lung cancer deaths, particularly rising among men but not in women, suggesting it was not entirely artefactual (Peto, 2023). Subsequently, research on a group of lung cancer patients and other patients was initiated, and it emerged that half of the other patients smoked cigarettes, while nearly all of the lung cancer patients did so (Peto, 2023).

Notably, the work on cancer causes in 1981 by renowned epidemiologists, Sir Richard Doll and Sir Richard Peto, began the notion that cancer may be an avoidable outcome of exposure to carcinogens (Doll and Peto, 1981, Peto, 2023). They quantified the fraction of cancers attributable to risk factors, with a focus on the largest contributors to cancer risk, which were tobacco, alcohol, diet, infections, and environmental carcinogens (Doll and Peto, 1981). With tobacco being the top contributor to lung cancer deaths, their work on US cancer mortality data notes that women who were non-smokers in the 1970s had similar lung cancer rates to women in the 1950s and earlier, before smoking rose amongst women and led to lung cancer (Doll and Peto, 1981). They further noted 'these low rates of cancer may even prevail amongst the population as a whole if it were not for tobacco' (Doll and Peto, 1981). Two prospective studies investigated the generation of women who smoked early in life and for decades and suggested their risk of death from disease was 50% higher than during the 1980s when

smoking was relatively lower in prevalence among women (Jha et al., 2013, Thun et al., 2013). Moreover, their work gave rise to the notice that 'women who smoke like men, die like men' (BMJ Research News, 2013, Jha *et al.*, 2013, Thun *et al.*, 2013). This notion subsequently had led to the understanding that most observed sex differences may be reflective of the differences in health-related and environmental exposures between the sexes and thus has predominantly helped form the *a priori* understanding of why sex differences in cancer risk are observed.

More recently, research on contributions of intrinsic and sex-related biological risk factors towards sex differences in cancer incidence has increased. Biological sex differentiation happens at conception, when an ovum is fertilised by a sperm cell carrying an X or a Y chromosome, resulting in an embryo carrying an XX or XY chromosomal complement (Arnold, 2017, Mauvais-Jarvis *et al.*, 2020). The X chromosome is around 3 times larger than the Y chromosome, and has around 800-900 protein coding genes, whereas the Y chromosome has around 50-100 protein coding genes (Dirican and Nelson, 2024, National Human Genome Research Institute, 2024). The *SRY* gene located on the Y chromosome initiates the testis development, leading to the lifelong male-specific sex effects of gonadal hormones (Arnold, 2017, Mauvais-Jarvis *et al.*, 2020). The X and Y genes share ancestry and have 17 homologous gene pairs ('gametologs'), including ZFY-ZFX and UTY-UTX, which exhibit subtle functional differences (Bellott *et al.*, 2014, Lavorando *et al.*, 2024). They have been implicated in the regulation of gene expression at the transcriptional, epigenetic, translational, and post-translational levels and are potentially linked with sex-biased disease susceptibility (Lavorando *et al.*, 2024, Owens *et al.*, 2024).

In female cells, the presence of two X chromosomes is balanced by random inactivation of one X chromosome in each cell to limit differences in gene dosage between the sexes (Mauvais-Jarvis *et al.*, 2020). A subset of genes on the X chromosome escape inactivation

resulting in biallelic expression, termed ‘escape from X inactivation’ (EXITS) genes. Some EXITS genes are understood to be tumour suppressor genes and have been proposed to contribute to lower risk of certain cancers, such as bladder cancer in women (Mauvais-Jarvis *et al.*, 2020, Owens *et al.*, 2024). Collectively, differences in X and Y chromosomes are the fundamental source of genetic sex differences between men and women that persist throughout life and operate independently of, though in interaction with, sex hormones (Mauvais-Jarvis *et al.*, 2020, Owens *et al.*, 2024).

Considerable research has examined the influence of sex hormones on cancer incidence in men and women. They are understood to play an important role in tumorigenesis and cancer susceptibility by acting as modifiers of cancer-related mechanisms, including the cell cycle (mitosis, apoptosis, growth promotion), cancer stem cell self-renewal, the tumour microenvironment, immune function, and metabolism, thereby also contributing to observed sex disparities (Figure 1.1) (Clocchiatti *et al.*, 2016, Li *et al.*, 2019, Lopes-Ramos *et al.*, 2020, Mauvais-Jarvis *et al.*, 2020, Roy *et al.*, 2025, Zhao *et al.*, 2024). However, direct comparisons of the effects of sex-related biological factors on sex differences in cancer incidence remain difficult to isolate. Breast cancer provides a useful illustration. Although male breast cancer is rare, accounting for approximately 1% of all breast cancer cases, elevated oestradiol levels play a comparable aetiological role in men as in postmenopausal women (Anderson *et al.*, 2010, Brinton *et al.*, 2015, Brinton *et al.*, 2008). A prospective case–control analysis in men reported relative risks on the order of two- to three-fold when comparing individuals in the highest versus lowest quartiles of circulating oestradiol levels, similar to risk for postmenopausal women (Brinton *et al.*, 2015). More broadly, the higher incidence of cancers at non-reproductive sites—including liver, colorectal, oesophageal and haematological malignancies—in men and postmenopausal women compared with premenopausal women suggests that oestrogen signalling may confer a degree of protection and thereby contribute to observed sex disparities (Clocchiatti *et al.*, 2016). Nevertheless, evidence from childhood

and adolescent cancers, which exhibit a male predominance despite minimal differences in circulating sex hormones, suggests that the role of hormones may be partial and modifying rather than wholly explanatory (Lopes-Ramos *et al.*, 2020, Williams *et al.*, 2019).

Height is an intrinsic, non-modifiable biomarker positively associated with cancer risk across several sites and populations (Green *et al.*, 2011, Gunnell *et al.*, 2001). Although the mechanisms are not well established, height has been hypothesized to be a proxy for cell count, as well as hormonal and growth factor exposures, and thus being taller may be associated with cancer risk through increased cell turnover mediated by growth factors, or through increased cell number (Fu *et al.*, 2020b, Green *et al.*, 2011, Walter *et al.*, 2013). As the majority of non-reproductive cancer sites have higher incidence in men compared to women, and men on average are taller than women, height has been suggested to partially mediate the sex differences in incidence (Fu *et al.*, 2020b, NHS Digital, 2024, Walter *et al.*, 2013). An analysis in the Health Professionals Follow-up study and Nurses' Health study reported 35% (95% confidence interval 24-46%) of the excess risk observed in men was mediated by height (Fu *et al.*, 2020a).

Collectively, there is broad consensus that a range of environmental, lifestyle, biological, hormonal, and genetic risk factors contribute to observed sex differences in cancer incidence. However, the relative contributions of these risk factors, and the extent to which they explain cancer risk differences between men and women, remain incompletely understood.

## 1.3 Epidemiological Concepts and Definitions

A risk factor is the cause of a disease when alterations in frequency or intensity of the factor, without concomitant alterations in other factors, are followed by changes in the frequency of occurrence of the disease, following a latent or induction period of time (Lagiou *et al.*, 2018). At an individual level, the exposure to a cause of disease implies their likelihood of developing

the disease is higher, although there is no certainty this will happen (Lagiou *et al.*, 2018). Causality is rarely characterised by a simple correspondence between an exposure and specific disease; instead, exposures are either determined as necessary (essential in aetiology of disease) or sufficient causes (a set of exposures that inevitably lead to disease) (Lagiou *et al.*, 2018). For most diseases, there is no one necessary cause and exposures are 'component' causes that have a contributory role within sufficient causal mechanisms (Lagiou *et al.*, 2018).

The complexity of biological systems and incomplete understanding of many of the underlying processes hinder a deterministic explanation of disease causation (Lagiou *et al.*, 2018). Therefore, the concept of causation of disease is understood in a probabilistic (stochastic) sense and evaluated using statistical methods, with the aim of identifying causes of diseases (Lagiou *et al.*, 2018). Observational epidemiological studies, however, are inherently vulnerable to systematic error, which may produce conflicting findings or misleading inferences and, in some cases, unwarranted concern or confusion regarding public health priorities (Lagiou *et al.*, 2018, Taubes, 1995).

Confounding is one major source of systematic error, generated when a risk factor causally or strongly associated with a disease, is also related to the exposure under investigation, without lying on the pathway between the exposure and disease (Goodman and Samet, 2017, Lagiou *et al.*, 2018). Studies address this by identifying or speculating confounders and by adequately conceptualizing and accurately measuring them (Goodman and Samet, 2017, Lagiou *et al.*, 2018). In practice, however, complete control of confounding is rarely achievable due to imperfect measurement, unmeasured variables, or limited data availability, resulting in residual confounding that remains (Goodman and Samet, 2017, Lagiou *et al.*, 2018).

Given these limitations, the assessment of causality is not applied entirely mechanistically or scored quantitatively; instead requires scientific judgement, synthesis of evidence, and

consideration of alternative explanations. Several criteria for inferring causation from epidemiological investigations have been determined (Hill, 1965, International Agency for Research on Cancer, 1987, Lagiou *et al.*, 2018, MacMahon *et al.*, 1960). One such example is the Bradford Hill (1965) criteria, a widely applied framework comprising nine considerations for inferring causal associations: strength, consistency, specificity, temporality, biological gradient, plausibility, coherence, experimental evidence, and analogy. As epidemiological evidence accumulates, the perceived likelihood of a causal association between a particular exposure and disease moves forward and backward on a continuous spectrum until a particular consensus of sufficiency is established (Lagiou *et al.*, 2018).

## 1.4 Central Aim and Analytical Approach

Although sex is sometimes described in epidemiological studies as a 'risk factor' for cancer incidence, the terminology requires qualification. Use of 'sex' as a risk factor may represent intrinsic determinants of biological sex, such as sex chromosomes and sex hormones. For cancers common to both sexes, however, sex may also represent different risk profiles with respect to extrinsic risk factors, e.g. a different likelihood of being exposed to a risk factor such as smoking or to a different level of a risk factors. Therefore, understanding the sex differences in cancer incidence requires consideration of mediation (Bours, 2021, Bours, 2023, Hernan and Robins, 2025). Mediation concerns the extent to which observed sex differences operate through intermediate exposures (Bours, 2021, Bours, 2023, Hernan and Robins, 2025). Assessing mediation using observational studies is challenging, as it relies on strong formal assumptions and requires careful consideration of relevant exposures, as well as of residual and unmeasured confounding (Bours, 2021, Bours, 2023, Hernan and Robins, 2025). A further consideration is the extent to which sex may be an effect modifier of the associations of established risk factors with cancer incidence. This conceptual framework underpins the objectives and analytical approach adopted in this thesis.

The central aim of this thesis was to conduct a comprehensive investigation of sex differences in cancer incidence and to assess the extent to which observed disparities persisted after accounting for established and measured risk factors. This was achieved through an outcome-wide analytical approach, allowing standardised analyses across multiple cancer sites within a cohort (UK Biobank). Using a consistent modelling framework enabled direct comparison of sex differences and their attenuation across cancers.

Cancer development is a multistage process influenced by complex interplay of environmental, health-related, and biological factors acting at different stages of carcinogenesis (Figure 1.1). Sex differences in cancer incidence may be indicative of the differences in risk factors at different stages of carcinogenesis. To organise and evaluate contributory risk factors within an epidemiological framework for the thesis analyses, risk factors are classified as extrinsic and intrinsic (Figure 1.2). Extrinsic risk factors include environmental and health-related exposures that may differ by sex and are potentially modifiable, whereas intrinsic factors encompass inherent biological and sex-related characteristics, as well as underlying medical conditions, that may influence cancer susceptibility. This classification is necessarily simplified and partially overlapping but is intended to guide as a structured framework for organising the range of risk factors evaluated throughout the thesis.

For example, body size, as measured by body mass index (BMI), is classified as an intrinsic biological factor reflecting differences in overall body size and composition between men and women (Figure 1.2). This is conceptually distinct from overweight and obesity, which are treated as health-related extrinsic factors, as they represent physiological states arising from the cumulative influence of health-related and environmental exposures. In contrast, underlying medical conditions and diseases are categorised as intrinsic risk factors, as they proxy intrinsic chronic biological or pathological states, thus influencing cancer

susceptibility. For example, inflammatory bowel disease is an immune-mediated condition that lies on the causal pathway for colorectal cancer risk but also arises partly from extrinsic exposures. Accordingly, this classification is intended to be analytic rather than absolute.

In the UK Biobank cohort, 28 cancer endpoints—including 15 main cancer sites and 13 subtypes—were selected for analyses, each with at least 100 incident cases, and collectively comprise the *pan-cancer* scope of the prospective analyses undertaken in this thesis. The selection of cancer risk factors evaluated in relation to sex differences was guided by an *a priori* rationale informed by an existing large-scale outcome-wide analysis in the NIH–AARP Diet and Health Study (Jackson *et al.*, 2022), the *Textbook of Cancer Epidemiology* (3rd edition) (Adami *et al.*, 2018), and classifications of agents with sufficient or limited evidence of carcinogenicity as defined by the International Agency for Research on Cancer (IARC) Monographs (Volumes 1–140) (International Agency for Research on Cancer, 2025b). This framework informed both the literature review (Chapter 2) and the selection of risk factors assessed for baseline, site-specific, and chapter-specific analyses.

## 1.5 Thesis Objectives and Structure

This thesis comprises three research projects exploring sex differences in cancer incidence. Including and leading up to those three chapters, the thesis is laid out in the following format:

Chapter 2 reviews the background literature on sex differences in cancer incidence. It examines temporal and age-specific patterns of cancer incidence by sex; reviews sex differences in the distribution and effects of extrinsic and intrinsic risk factors (as organized in Figure 1.2) hypothesized to contribute to disparities in cancer incidence; and summarizes sex differences by cancer site, along with the risk factors hypothesized to underlie these differences.

Chapter 3 describes the data sources and methodology for the thesis. It details the UK Biobank cohort data (including how the data were obtained and structured for analyses) and explains the primary statistical approach used throughout the research (Cox proportional hazards regression modelling with attained age as the underlying time scale variable). In addition, chapter-specific statistical analyses are described within each subsequent research chapter.

Chapter 4 presents the first research project, which is a pan-cancer analysis of sex differences in incidence of 15 cancers (and 13 subtypes) common to men and women. Here, the extent to which the observed sex differences are attenuated when adjusting for established extrinsic and intrinsic risk factors is evaluated.

Chapter 5 builds on the pan-cancer findings by examining whether sex differences in adiposity measures (beyond BMI) contribute to the higher incidence of oesophageal adenocarcinoma, kidney cancer, and rectal cancer in men compared to women.

Chapter 6 investigates the higher incidence of lung adenocarcinoma in women among never-smokers. Contributions of risk factors that differed in prevalence by sex (household fuel use and asthma) and sex-specific factors (menopausal status) were evaluated.

Chapter 7 discusses and summarises findings of the thesis, highlights the strengths and limitations of the work, and reviews future research directions.

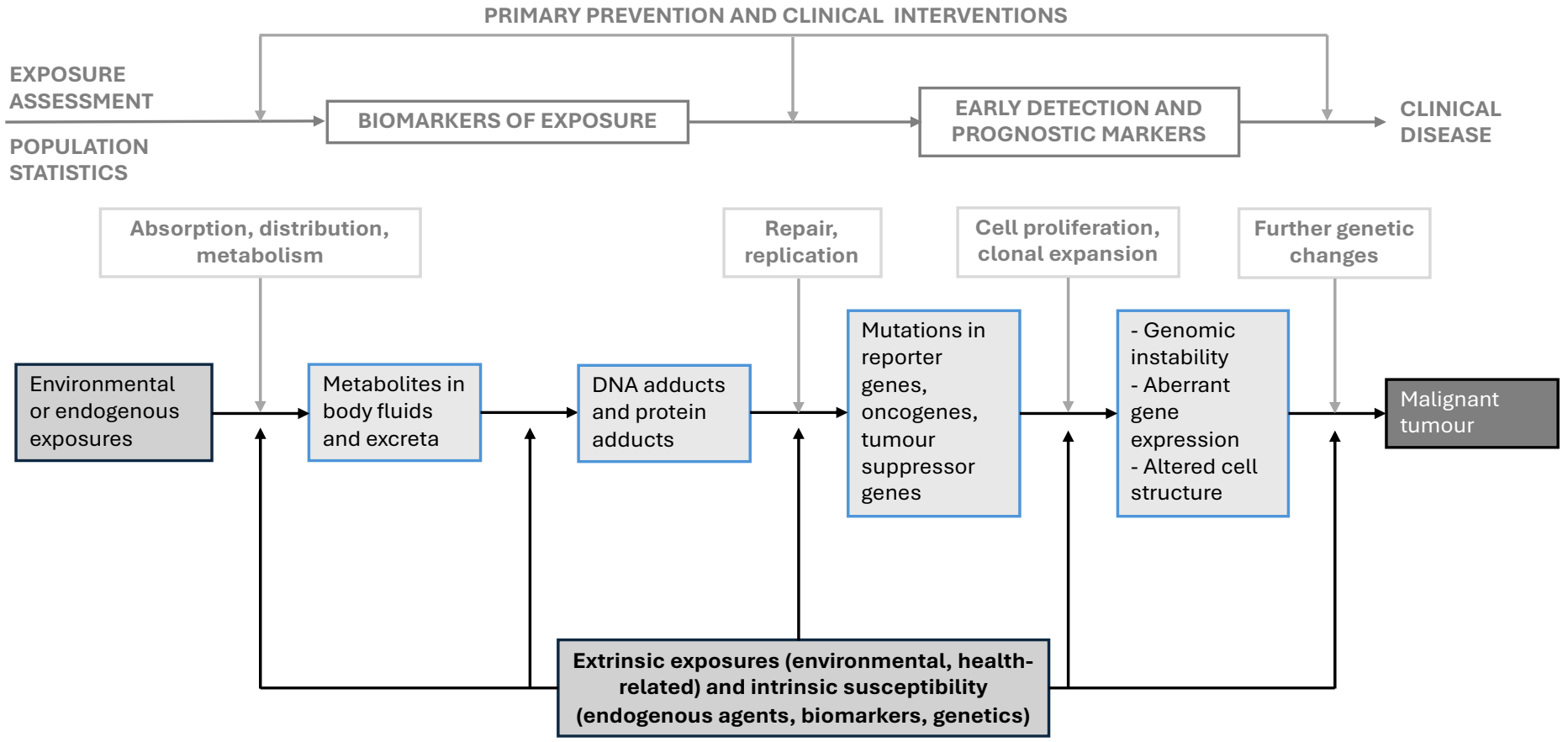



**Figure 1.1 Conceptual representation of the contributions of extrinsic and intrinsic risk factors across stages of carcinogenesis.**

Environmental, health-related, and endogenous exposures may influence cancer development through processes including absorption and metabolism, DNA damage, mutation, clonal expansion, and genomic instability, ultimately leading to malignant transformation. Intrinsic susceptibility—including genetic factors, biological biomarkers, and endogenous agents—may modify cancer risk at multiple stages of this pathway.

Opportunities for primary prevention, early detection, and clinical intervention exist at multiple points along this pathway. *Adapted from Boffetta and Mucci (2018), based on the National Research Council (1987).*

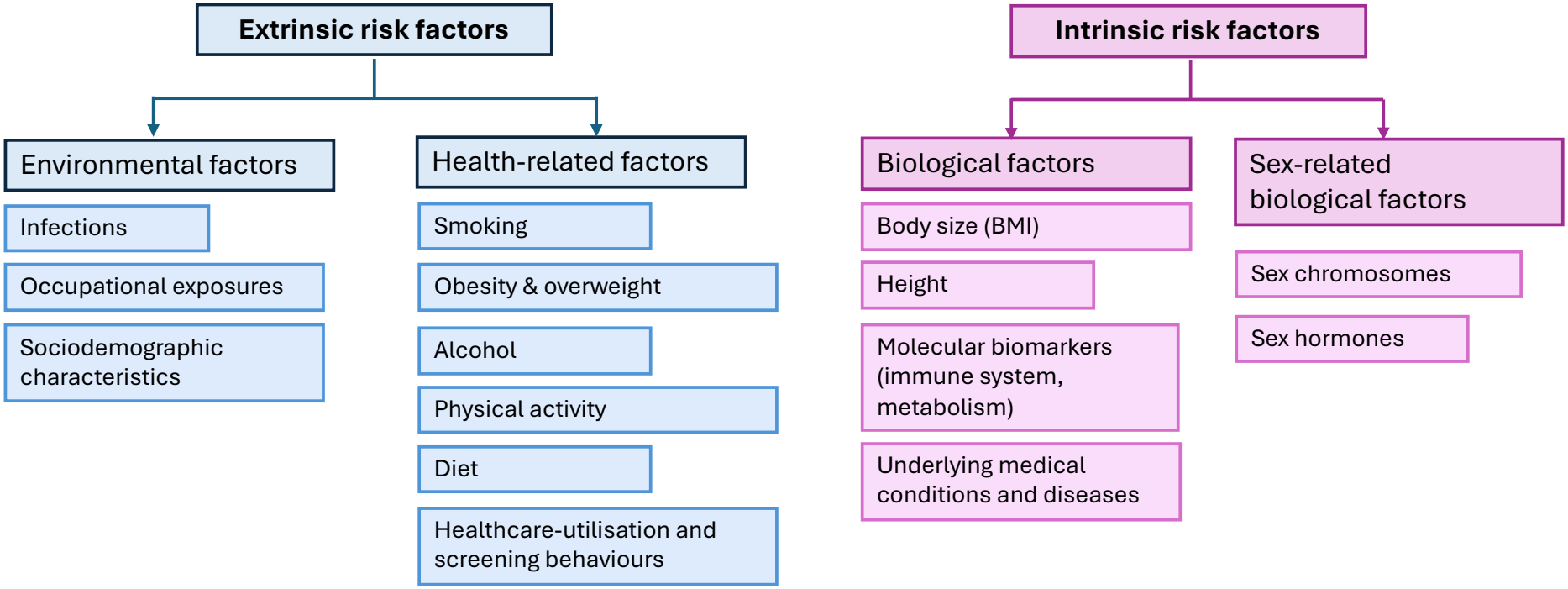



**Figure 1.2 Classification of extrinsic and intrinsic risk factors in epidemiology hypothesized to contribute to sex differences in cancer incidence.**

Extrinsic factors include environmental and health-related exposures that differ by sex and are potentially modifiable, whereas intrinsic factors reflect biological and sex-related characteristics that may influence cancer susceptibility. This classification is necessarily simplified and partially overlapping but is intended to provide a structured framework for organising the range of risk factors evaluated in this thesis.

# Chapter 2 A Review of the Sex Differences in Cancer Incidence

Here, the literature on sex differences in cancer incidence is reviewed, structured by the temporal trends and age-related trends observed in cancer incidence between men and women, sex differences across key extrinsic and intrinsic risk factors, and by cancer sites. The classification of risk factors as extrinsic (environmental and health-related exposures that may differ by sex and are potentially modifiable) and intrinsic (biological and sex-related characteristics that may influence cancer susceptibility) has been adopted as a simplified and structured framework for organising risk factors evaluated throughout the thesis (Figure 1.2).

## 2.1 Trends for Sex Differences in Cancer Incidence

### 2.1.1 Temporal Trends

The magnitude of sex differences in cancer incidence in the UK has decreased overall in the last four decades, as cancer incidence rates have declined and stabilised in men, while a modest annual increase has been observed in women (Figure 2.1) (Ervik *et al.*, 2024, Shelton *et al.*, 2024). However, the incidence of cancers altogether continues to be significantly higher in men compared to women (Ervik *et al.*, 2024, Shelton *et al.*, 2024). In 2017, the overall age-standardised incidence rates (ASIRs) for cancers common to men and women (excluding breast) aged 35 and above in the UK (England) were 473.9 per 100,000 for men and 314.5 per 100,000 for women (Ervik *et al.*, 2024). In comparison, in 1971, the overall ASIRs were 486.9

per 100,000 for men and 198.9 per 100,000 cases for women (Ervik *et al.*, 2024). The ASIRs are for multiple cancer sites combined, which are lip, oral cavity and pharynx (ICD-10 codes C00–C14), oesophagus (C15), stomach (C16), liver (C22), gallbladder (C23), pancreas (C25), larynx (C32), lung (C34), colorectum (C18–C20), melanoma (C43), Kaposi sarcoma (C46), thyroid (C73), kidney (C64), bladder (C67), brain and central nervous system (C70–C72), Hodgkin lymphoma (C81), non-Hodgkin lymphoma (C82–C85), and multiple myeloma (C90), and leukaemia (C91–C95) (Figure 2.1).

These trends in cancer incidence overall are largely or partially attributable to the trends in cancer risk factors (Brown *et al.*, 2018, Shelton *et al.*, 2024). Tobacco smoking and overweight and obesity are two risk factors with the highest population attributable fractions for most cancers (Brown *et al.*, 2018). Population attributable fractions represent the proportion of disease cases in a population that are attributed to exposure to a specific risk factor, based on the assumption that the association between the risk factor and the disease is causal (Brown *et al.*, 2018). Tobacco smoking contributes to the largest proportion of all cancer incidence in the UK at 15.1% (Brown *et al.*, 2018). Nonetheless, tobacco smoking has declined in prevalence over the last 25 years, predominantly in men, with legislative public health initiatives and mass media efforts (Brown *et al.*, 2018, Shelton *et al.*, 2024). Overweight and obesity contributes to the second highest proportion of all cancer incidence in the UK at 6.3%, for which the prevalence is rising for both men and women (Brown *et al.*, 2018, House of Commons Library, 2025). Between 1993 and 2022, the proportion of individuals in the UK who were obese increased from 14.9% to 28.9%, and from 52.9% to 64.3% who were overweight to obese (House of Commons Library, 2025). The increases in people who are overweight and obese have been the largest between 1993-2001 and since then have been more gradual (NHS Digital, 2022b). More specifically, obesity increased from 13.0% to 27.0% in men and 16.0% to 29.0% in women from 1993 to 2019 (NHS Digital, 2022b). Overweight including obesity increased from 58.0% to 68.0% in men and 49.0% to 60.0% in women from

1993 to 2019 (NHS Digital, 2022b). Between 2022-2023, the prevalence of being overweight including obese was higher amongst men compared to women, with 69.2% men and 58.6% women classified in both categories (Office for Health Improvement & Disparities, 2024).

Given that nearly 40% of cancer incidence in the UK is attributable to known modifiable risk factors (Brown *et al.*, 2018), and that temporal trends in cancer incidence differ by sex in ways that broadly mirror sex-specific patterns of risk factor exposure, it is reasonable to hypothesise that differential risk factor distributions contribute substantially to observed sex differences in cancer incidence.

### 2.1.2 Age-related Trends

Among individuals aged 35 and older, cancer incidence is consistently higher among men compared to women for cancers common to both sexes (except breast cancer), with the magnitude of this sex difference increasing with age (Figure 2.2) (Cook *et al.*, 2009, Edgren *et al.*, 2012). ASIRs per 100,000 between the periods of 2013-2017 in the UK (England) were 57.9 for men and 53.7 for women from ages 33-39, and 2434.8 for men and 1375.8 for women aged 85 and above (Ervik *et al.*, 2024). Age-related increases in sex differences in cancer incidence have been described previously in cross-sectional epidemiological studies that have used national level data and reported male-to-female incidence rate ratios (MF IRRs) (Cook *et al.*, 2009, Dong *et al.*, 2020, Edgren *et al.*, 2012). One such study analysed data from the 'Cancer Incidence in Five Continents' database for the period 1998-2002 (Edgren *et al.*, 2012). Across 35 cancer sites examined, incidence at 32 sites was higher in men compared to women, with over four-fold higher incidence at five sites (cancers of the larynx, hypopharynx, lip, and urinary bladder, and Kaposi sarcoma) and twofold incidence at 15 sites.

When stratified by age groups for individuals aged above 40 (40-59 years, 60-74 years, and ≥75 years), MF IRRs for 14 sites increased with age, which were cancers of the colon, eye, larynx, lung, melanoma of the skin, myeloid myeloma, myeloid leukaemia, other endocrine, rectum,

salivary glands, small intestine, and stomach, and urinary bladder (Edgren *et al.*, 2012). Eight cancer sites had stable MF IRRs across age groups, which were cancers of the hypopharynx, nasopharynx, other urinary organs, non-Hodgkin lymphoma, lymphoid leukaemia, bone, and brain and the nervous system, and kidney. 10 cancer sites had declining MF IRRs with age, which were Kaposi sarcoma, cancers of the lip, oesophagus, tonsil, liver, tongue, nose, oral cavity, and pancreas, and Hodgkin lymphoma. Three cancer sites—anus, gallbladder, and thyroid—had higher cancer incidence in women compared to men, and for all three sites MF IRRs attenuated towards the null with age. Findings from US-based studies SEER data reported broadly similar age-related patterns in MF IRRs across cancer sites to those observed by Edgren *et al.* (2012) (Cook *et al.*, 2009, Dong *et al.*, 2020).

Sex differences in cancer incidence have also been observed in children and young adults (aged 0-19 years) (Edgren *et al.*, 2012, National Disease Registration Service, 2021). Childhood cancers account for less than 1% of all cancer diagnoses in the UK, with incidence higher in boys than girls, who account for approximately 55% and 45% of cases respectively (Edgren *et al.*, 2012, National Disease Registration Service, 2021). Among young adults, cancer incidence is broadly similar between boys and girls for most sites, although higher incidence in boys has been reported for sites including melanoma of the skin, gallbladder, thyroid, and salivary glands (Edgren *et al.*, 2012).

Age-related changes in the magnitude of sex differences in cancer incidence may reflect the cumulative contributions of sex-specific biological and lifestyle-related risk factors across the life course.

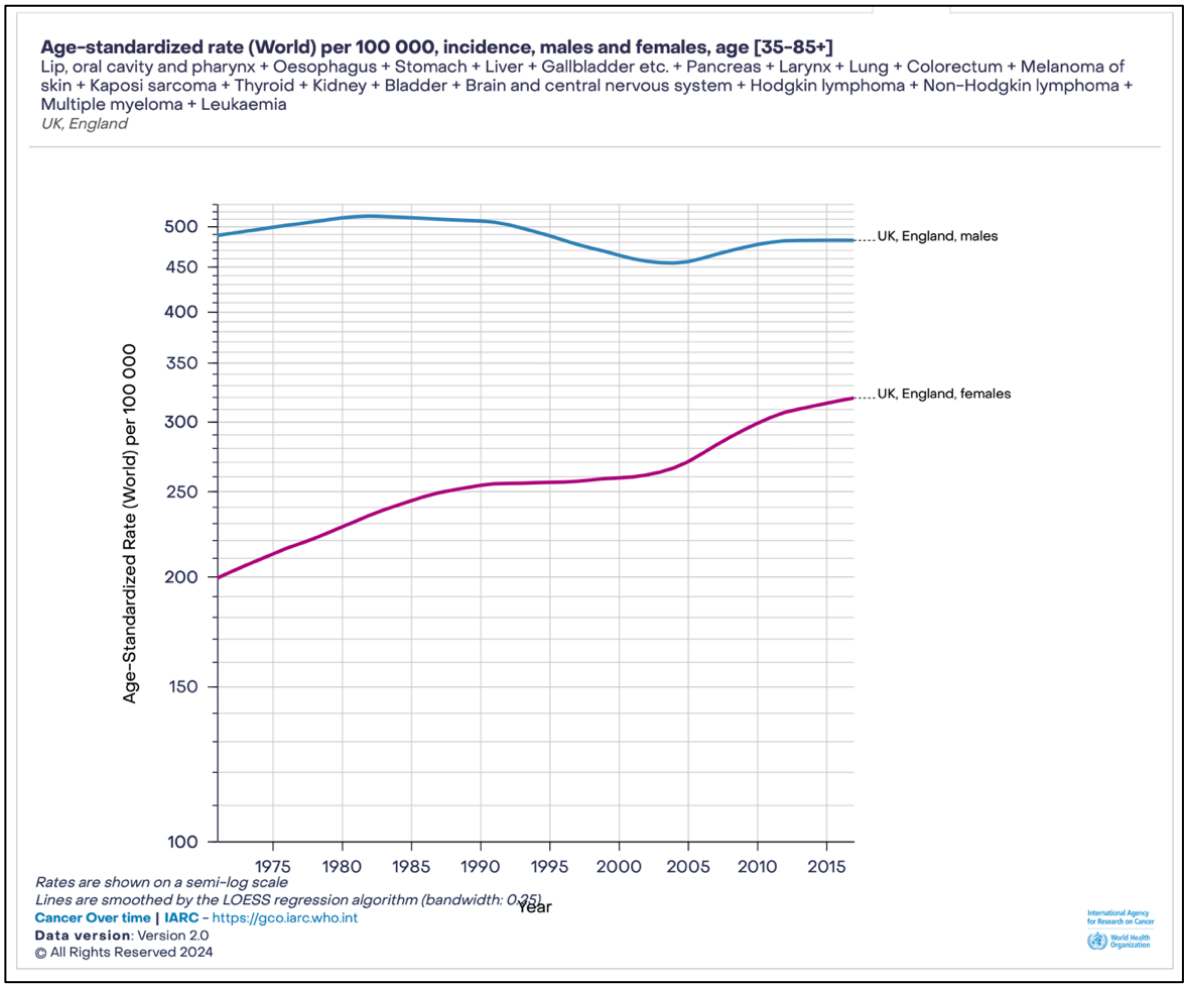



**Figure 2.1 Age standardised incidence rates per 100,000 from 1971-2017 for cancers common to both men and women aged 35 and above in the UK (England).**

*Source: Global Cancer Observatory; cancer incidence data are obtained from the England Cancer Registry and available for the period 1971–2017 (Ervik et al., 2024).*

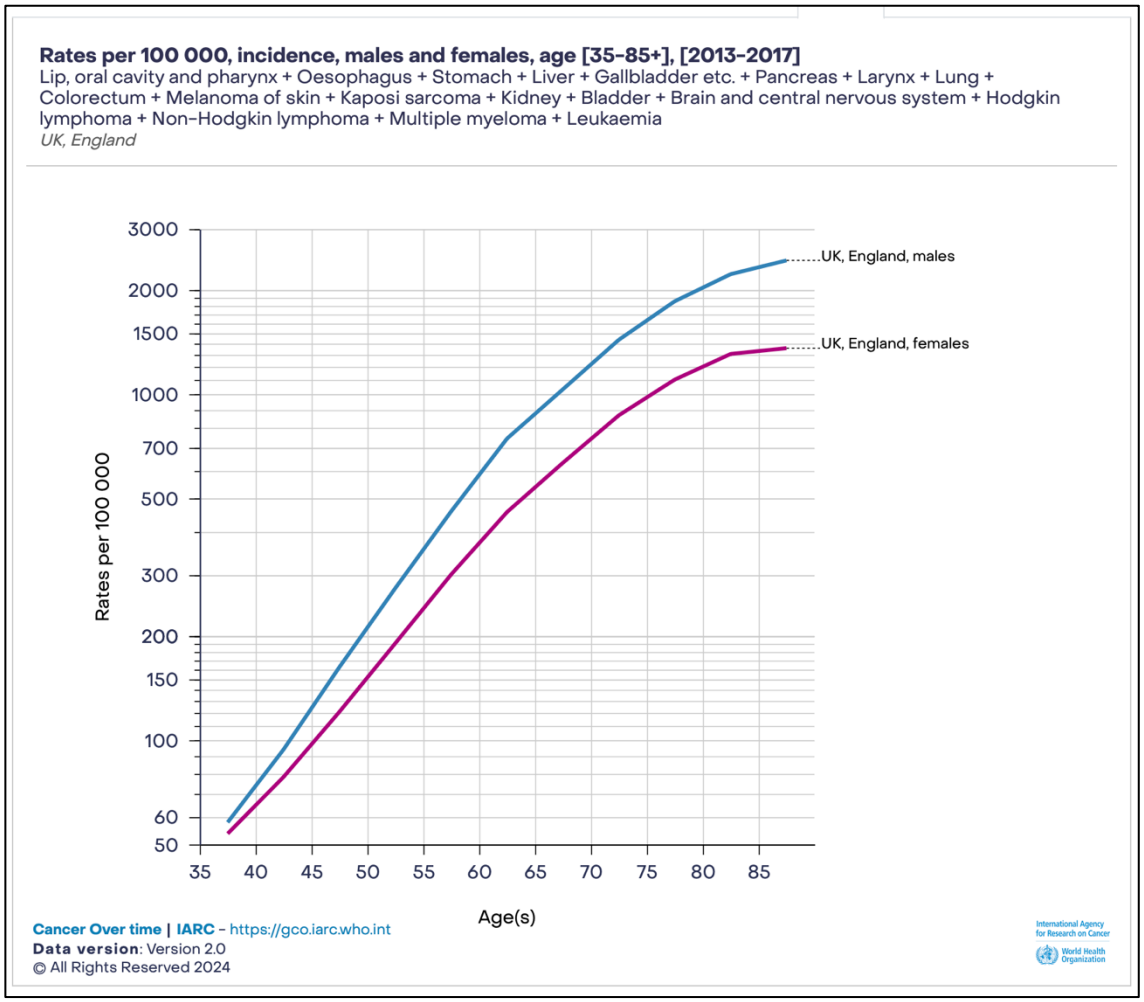



**Figure 2.2 Cancer incidence rates per 100,000 from 2013-2017 for cancers common to both men and women aged 35 and above in the UK (England).**

*Source: Global Cancer Observatory; cancer incidence data are obtained from the England Cancer Registry (Ervik et al., 2024).*

## 2.2 Sex Differences Across Extrinsic Risk Factors

### 2.2.1 Sociodemographic Characteristics

Patterns of sex differences in cancer incidence across sociodemographic characteristics, such as socioeconomic deprivation, educational attainment, and ethnicity, are assessed to infer whether observed sex disparities may be driven predominantly by intrinsic biological factors—if sex differences remain consistent across groups—or by extrinsic environmental and health-related factors if they vary between groups (Jackson *et al.*, 2022, Scelo *et al.*, 2018, Tosakoon *et al.*, 2024). Cancer incidence rates vary across sociodemographic characteristics and cancer sites, and there isn't a consistent pattern or trend for all sites (Chen *et al.*, 2025, Hiatt *et al.*, 2025, Hodge *et al.*, 2023, Malcomson *et al.*, 2024).

#### Socioeconomic Deprivation

Cancers of the lung and stomach have the highest incidence in areas with highest deprivation levels in the UK (England and Wales), with lung cancer being approximately threefold higher in the most deprived areas (Hiatt *et al.*, 2025, NHS Digital, 2021). Whereas cancers of the breast and melanoma (and even prostate and non-melanoma skin cancer) have reverse gradients, with higher incidence in least deprived areas (Hiatt *et al.*, 2025, NHS Digital, 2021).

Across all cancer sites combined, ASIRs are higher for men in least deprived areas compared to women in the most deprived areas (649 versus 621 per 100,000 people) (Hiatt *et al.*, 2025, NHS Digital, 2021). The ASIRs for men were 756 per 100,000 people in the most deprived areas and 649 in the least deprived areas. For women, the corresponding rates were 621 per 100,000 people in the most deprived areas and 519 in the least deprived areas (NHS Digital, 2021). Although men have higher baseline cancer incidence, socioeconomic deprivation increases cancer incidence proportionally in both men and women.

## Educational Attainment (Qualifications)

Trends in cancer incidence rates by educational attainment for men and women are broadly similar to those observed for socioeconomic deprivation, and educational attainment is commonly used as a measure of socioeconomic status and a social indicator of health (Hodge *et al.*, 2023). Although correlated with socioeconomic deprivation, educational attainment is also independently associated with cancer risk and may be closely related to cancer screening uptake (Chadeau-Hyam *et al.*, 2020, Elbasheer, 2024, Hodge *et al.*, 2023). A prospective cohort study within the American Cancer Society's Cancer Prevention Study-II cohort, comprising over 148,000 participants (52.5% women), examined associations between educational attainment and site-specific cancer risk while adjusting for multiple cancer risk factors (Hodge *et al.*, 2023). The study reported inverse associations between educational attainment and cancer risk in men for cancers of the head and neck, oesophagus, stomach, colon, rectum, liver, lung, kidney, and bladder, and positive associations for melanoma and prostate cancer. In contrast, among women, associations were inverse for colon and lung cancer, null for oesophageal, rectal, liver, kidney, stomach, and bladder cancers, and positive for breast cancer and melanoma. Their findings were consistent with previous literature, suggesting that educational attainment contributes to the effects of other health-related behaviours on cancer risk, particularly smoking, cancer screening behaviours, and, among women, reproductive factors (Hodge *et al.*, 2023).

## Ethnicity

The term 'ethnicity' can have different meanings in different contexts; here, it is considered as a socially patterned and multidimensional characteristic rather than as a fixed biological exposure. Ethnicity is also self-defined and context-dependent, and requires consideration of a multitude of relevant factors, including racism and discrimination, socioeconomic status, social class, environmental exposures, health beliefs and practice, education, diet and nutrition, religion, migration history, ancestry, and cultural and lived experience (Kaplan

and Bennett, 2003, Lam *et al.*, 2023). Therefore, broad ethnic categories may obscure substantial heterogeneity within groups (Kaplan and Bennett, 2003, Lam *et al.*, 2023). In observational studies, interpretation of ethnicity is further complicated by variation in how ethnicity is collected, recorded, categorised, and analysed, as well as by incomplete recording, potential misclassification, under-representation of minoritised ethnic groups, and aggregation into broad categories (Lam *et al.*, 2023).

Additionally, disentangling sex and gender from race and ethnicity is difficult due to intersectionality. Intersectionality posits that social positions exist on a hierarchy of social power that are not independent, and shape human behaviours and experiences jointly (Bauer *et al.*, 2021, Crenshaw, 1989). Ethnicity may intersect with sex, socioeconomic deprivation, migration status, occupation, geography, and healthcare access in ways that shape cancer-relevant exposures and opportunities for prevention or diagnosis (Bauer *et al.*, 2021, Crenshaw, 1989). Single-axis analyses, such as comparing sexes overall or ethnic groups overall, may therefore mask heterogeneity among subgroups.

Few epidemiological studies have made direct comparisons assessing contributions towards sex differences in cancer incidence by ethnicity (Benigni, 2007, Tosakoon *et al.*, 2024). A cross-sectional study using the US SEER program, a national registry database covering 47.9% of the population, between 2000-2019 explored sex differences across ethnicities for 49 cancer sites in adults aged over 20 years (Tosakoon *et al.*, 2024). They reported 44 cancer sites had higher risk in men compared to women (MF IRR >1), and for seven sites, MF IRRs varied by race and ethnic group. These were cancers of lip, tongue, hypopharynx, retroperitoneum, larynx and pleura, and Kaposi sarcoma. The authors hypothesized the variations in MF IRRs for cancers of the oral cavity were related to patterns of smoking, alcohol, and occupational exposures. For most other cancer sites, the authors concluded the low variation in MF IRRs by race and ethnicity were suggestive of sex differences in biological

risk factors (genetic factors or immune function) influencing cancer susceptibility (Tosakoon *et al.*, 2024). However, this interpretation should be considered cautiously, as broad race and ethnicity categories do not directly distinguish biological mechanisms from differences in social, environmental, behavioural, or healthcare-related exposures.

### 2.2.2 Smoking

Tobacco use remains the leading risk factor for cancer incidence and mortality worldwide (World Health Organization, 2021). In 2020, an estimated 22.3% of the global population (7.8% of women and 36.7% of men) used tobacco (World Health Organization, 2021). In the UK, 11.9% of adults (13.7% men and 10.1% women) aged 18 or over (6.0 million individuals) were identified as current smokers in 2023 (Office of National Statistics, 2024). Smoking prevalence in the UK has been declining since the 1970s due to measures such as increased taxation on tobacco products, advertising bans, the implementation of smokefree policies, and public education efforts promoting cessation (Brown *et al.*, 2018). For comparison, in 2005, 12.4% of women and 17.7% of men in the UK were current smokers (Brown *et al.*, 2018).

Smoking gained popularity in the UK during the first half of the 20$^{th}$ century, first amongst men and later amongst women (Peto *et al.*, 2000). Although women historically smoked less than men, the decline in smoking prevalence from the 1950s was slower and slightly delayed for women than for men (Brown *et al.*, 2018, Peters *et al.*, 2014, Peto *et al.*, 2000). Furthermore, as smoking habits among women changed, i.e., increases in smoking intensity and earlier initiation ages, the prevalence patterns started to become comparable to those in men (Peto *et al.*, 2000). In 2010, current smokers in England were estimated to consume an average of 12.7 cigarettes a day, with men smoking 13.3 and women smoking 12.1 cigarettes a day (Health and Social Care Information Centre, 2013). Although sex differences in smoking prevalence and behaviours persist, in the UK, they have become considerably narrower over time (Peters *et al.*, 2014).

Smoking was estimated to be attributable to 15.1% of all cancer cases (17.7% for men and 12.4% for women) in the UK in 2015 (Brown *et al.*, 2018). Smoking is a causal risk factor for cancers of the oral cavity, oesophagus, stomach, colorectum, liver, pancreas, lung, breast, kidney, bladder, and some types of leukaemia (IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 2012). Some studies have hypothesized that smoking may have differential effects on cancer risk by sex, and that women who smoke at levels comparable to men may have a higher risk of certain cancers, particularly lung cancer (Freedman *et al.*, 2008, Mederos *et al.*, 2020, Peters *et al.*, 2014). However, evidence from prospective cohort studies report no consistent difference in smoking-related cancer risk between men and women (Freedman *et al.*, 2008, Thun *et al.*, 2013). With respect to self-reported smoking by sex, it has been studied have generally high concordance with biochemical measures in population surveys, and under-reporting has only been noted in settings with social desirability pressure (for example, pregnant women) (Jackson *et al.*, 2019, Shipton *et al.*, 2009, Woodward *et al.*, 1991).

Jackson *et al.* (2022) in their investigation of sex differences in cancer incidence across 21 cancer sites adjusted for prevalent smoking behaviours and assessed effect modification by smoking status. They noted sex differences remained consistent for most cancer sites and none of the interactions were significant (Bonferroni-corrected $p>0.002$) for any cancer sites. However, the authors noted that their analyses could not exclude the influence of unmeasured or residual confounding from smoking.

### 2.2.3 Alcohol

Men typically consume more alcohol than women, both in terms of frequency and volume (Figure 2.3) (Brown *et al.*, 2018, Stiebahl, 2024). In 2019, an estimated 2.6 million deaths globally were attributed to alcohol use, of which 2 million were in men (World Health Organization, 2024). According to the National Health Survey, 56% of the UK population

reported drinking alcohol in the previous week, with higher rates among men (61% compared to 51% for women) (Stiebahl, 2024).

Alcohol consumption in the UK is commonly measured in units, with one unit defined as 10 ml or 8 grams of pure alcohol—the approximate amount the body can metabolize in about an hour (National Health Services, 2024). In 2016, updated guidelines recommended a maximum weekly intake of 14 units for all adults, advising that consumption be spread over at least three days to mitigate health risks (Stiebahl, 2024). Previously, in 1995, guidelines had suggested daily limits of 3–4 units for men and 2–3 units for women (Stiebahl, 2024).

Alcohol use has been linked to between 2–4% of all cancer cases in the UK, although this may be an underestimation (Brown *et al.*, 2018). Patterns of alcohol consumption vary throughout the life course. In the UK, consumption typically rises during adolescence, plateaus in mid-life, and declines in older age (Britton *et al.*, 2015). However, the timing and extent of these changes differ by sex. Men tend to maintain higher consumption levels until their 70s, whereas women's average consumption decreases more significantly during early adulthood (ages 15–29) and again between 40–49 years (Britton *et al.*, 2015). Additionally, discrepancies between self-reported alcohol use in national surveys and sales data suggest that consumption levels are generally underestimated (Britton *et al.*, 2015). Heavy alcohol drinkers are also more likely to smoke, which introduces residual confounding when examining the relationship between alcohol and cancer risk (Bagnardi *et al.*, 2011, Bagnardi *et al.*, 2015). Nonetheless, alcohol has been causally linked to increased risks for several types of cancer, including, but not limited to, cancers of the oral cavity, oesophagus, breast, colorectum, liver, kidney, pancreas, and melanoma (World Health Organization, 2024).




Notes: The Health Survey for England was suspended in 2020 due to the coronavirus pandemic. The dotted area of the chart highlights this gap in the data series.

Source: Health Survey for England 2022, Adult drinking

**Figure 2.3 Alcohol consumption levels in the UK for men and women (2012-2022).**

*Data reported as per the Health Survey for England, 2024. Figure sourced from Stiebahl, 2024.*

## 2.2.4 Infections

In the UK, carcinogenic infections are calculated to contribute to 3.6% of all cancer cases (3.1% in men and 4.2% in women) (Brown *et al.*, 2018). The population attributable fractions from infections are higher for women due to sex-specific cancers (particularly cancer of the cervix, for which it is presumed oncogenic HPV infection is a necessary cause) (Brown *et al.*, 2018, Parkin, 2011).

High-risk HPV types are also implicated in anogenital and oral cavity cancers, and HPV is estimated to drive over 70% of oropharyngeal cancers, predominantly through HPV16 (Brenner *et al.*, 2020, Gillison *et al.*, 2015). In high-income countries, including the UK, HPV-driven oral cavity cancer incidence has risen particularly among men, despite similar HPV16 seroprevalence observed between sexes in population studies (Brenner *et al.*, 2020). This suggests that additionally sex differences in exposure patterns, sexual behaviour, vaccination uptake, or progression from infection to malignancy may contribute to the male excess in oral cavity cancer incidence. In contrast, declining smoking-related lung and oral cavity cancers

in men, alongside more modest declines or stable rates in women, indicate that changing distributions of both HPV infection and tobacco exposure may differentially shape cancer trends by sex (Gillison *et al.*, 2015).

To some extent, sex differences in liver cancer have also been hypothesized to be due to infections (Brown *et al.*, 2018), as there are sex differences in the seroprevalence of hepatitis B infections in the UK. According to the UK Health Security Agency, the estimated prevalence of hepatitis B surface antigen in the population in 2022 was 0.41% (95% CI 0.36-0.46) in women and 0.77% (95% CI 0.64-0.92) in men (Leeman *et al.*, 2024). Additionally, there are sex differences in the immunological responses mounted to chronic hepatitis infections, which may further contribute to the disparity (Klein and Flanagan, 2016, Steeg and Klein, 2016). Together, these patterns suggest that sex-specific distributions of carcinogenic infections, alongside differential exposure and biological response, may partially contribute to sex differences in oral, lung, and liver cancer incidence.

### 2.2.5 Occupations

Men have historically experienced greater occupational exposure to established carcinogens, and although changes in labour market participation have increased women's exposure to other occupational and environmental risks, overall exposure remains higher in men. The British occupational cancer burden study estimated that approximately 5.7% of cancer registrations in men and 2.1% in women were attributable to occupational carcinogens (Rushton *et al.*, 2012). The effect on risk from occupational carcinogens varies by cancer sites, with oral cavity, oesophagus, mesothelioma, lung, melanoma, and breast, and bladder cancers being of note (Rushton *et al.*, 2012).

Rushton *et al.* (2012) noted the population attributable fractions from occupational exposures to carcinogens in the UK may be 21.1% for lung cancer incidence in men and 3.3% in women. Similarly it was estimated that 7.1% of bladder cancer incidence registrations in

men and 1.9% in women may be attributable to occupations (UK), although some studies noted estimates may be even higher and range from 16% to 24% (Kirkali *et al.*, 2005, Rushton *et al.*, 2012, Vineis and Simonato, 1991). The carcinogens of importance are mainly asbestos (particularly important risk factor for mesothelioma), mineral oils, solar radiation, silica, diesel engine exhaust, coal tars and pitches, dioxins, environmental tobacco smoke, radon, tetrachloroethylene, arsenic and strong inorganic mists, as well as occupational circumstances such as shift work (relevant for breast cancer) and occupations as painters or welders (Cancer Research UK, 2023, Rushton *et al.*, 2012).

## 2.3 Sex Differences Across Intrinsic Risk Factors

### 2.3.1 Height

Height is considered an important but not the sole contributor towards sex differences in cancer incidence, as it has been studied to be positively associated with cancer risk across several sites and populations (Green *et al.*, 2011, Gunnell *et al.*, 2001). Adult height is determined by genetic and environmental influences till the first 20 years of life, and influenced by genetic, nutritional, and environmental risk factors that affect growth in utero and during childhood and adolescence (Green *et al.*, 2011). Hormone levels, especially of growth factors such as insulin-like growth factors, both in childhood and adult life are deemed relevant (Green *et al.*, 2011). Adult height is generally maintained until the beginning of the fifth decade, at which point it starts to decline (may decline by up to 2–4 cm on average) with to age-related changes (Fernihough and Mcgovern, 2015). Although the mechanisms are incompletely understood, being taller is considered a proxy for increased cell turnover mediated by growth factors, or for increased cell number, thus influencing both the number of cells at risk of acquiring somatic mutations and the rate of tumour proliferation (Fu *et al.*, 2020b, Green *et al.*, 2011, Gunnell *et al.*, 2001, Walter *et al.*, 2013).

According to the health survey for England in 2021, the average heights of men and women aged 35-44 years were 177.3 cm and 164.3 cm, respectively (NHS Digital, 2022a). Previous research on sex differences in cancer incidence has suggested height may explain part of sex differences (Fu *et al.*, 2020b, Kabat *et al.*, 2014, Walter *et al.*, 2013). To our knowledge, two prospective cohort studies have been used to assess the role of height in explaining sex differences in cancer incidence. Walter *et al.* (2013) used data from the US Vitamins and Lifestyle cohort study to explore risk of cancer incidence in men compared to women using Cox proportional hazards models, adjusting for age, race, and cancer risk factors

hypothesized to explaining some of the sex differences. With an average of 7.3 (SD 2.1) years of follow up, 3,466 participants had incident cancers in a cohort of 65,038 (50.6% women) participants with a mean age at recruitment of 61.5 (SD 7.4 years). They suggested height and height squared (non-linear height) explained up to 33% of excess risk in men at all cancer sites shared between men and women, and specifically the proportion of excess risk mediated by height may range from 90.9% (for kidney cancer), 57.3% (for malignant melanoma), and 49.6% (for haematological malignancies).

Fu *et al.* (2020b) analysed data from two larger prospective cohort studies based in the United States (Health Professionals Follow-up Study and Nurses' Health Study), with a total of 164,984 participants (70.1% women) who had 34 years of follow-up in total and estimated that height mediated up to 35% of the association at all shared sites, and more specifically 87% for gastrointestinal cancer sites, 45% for melanoma, 25% for kidney, 84% for brain, and 24% for haematological malignancies.

Although the results from both the studies suggest height may explain up to and over 30% of the excess risk of cancer incidence in men compared to women, their results have discrepancies when comparing site-specific excess risk mediated by height. Analyses by Walter *et al.* (2013) had relatively modest statistical power, and analyses by Fu *et al.* (2020b) may be influenced by differences in cohort populations and exposure distributions. Collectively, both studies suggest height contributes meaningfully but not entirely to sex differences in cancer incidence.

### 2.3.2 Body Mass Index

Body mass index (BMI) is a measure of body size and fat based on height and weight. The World Health Organization classifies people with a BMI <25kg/$m^2$ as not overweight nor obese (here, throughout the thesis referred to as 'normal BMI'), ≥25 to <30kg/$m^2$ as overweight and ≥30kg/$m^2$ as obese (NHS Digital, 2022a). While the average BMI in men and women is roughly

similar, the prevalence of people who are both overweight and obese is higher in men than in women in the UK (Table 2.1) (Brown *et al.*, 2018, NHS Digital, 2022a). After smoking, obesity and overweight are the second leading contributors to cancer incidence in the UK (6.3% of all cancer cases attributable to excess adiposity), and in some countries, they may even surpass smoking as the primary contributor (Brown *et al.*, 2018, Rask-Andersen *et al.*, 2023). For both men and women, obesity and overweight are associated with increased risk for cancers of the colorectum, breast (post-menopausal women), liver, kidney, pancreas, gastric cardia, oesophagus (adenocarcinoma), gallbladder, thyroid, multiple myeloma, and meningioma (International Agency for Research on Cancer, 2017). However, the overall population-attributable fractions for cancers associated with elevated BMI are higher in women than in men (5.4% vs 1.9%), largely reflecting contributions to sex-specific cancers such as endometrial and post-menopausal breast cancer (Renehan and Soerjomataram, 2016). For cancers common to both men and women (excluding breast cancer), elevated BMI is associated with risk increases of similar magnitude in men and women (Lauby-Secretan *et al.*, 2016).

Nonetheless, there are also sex differences in body composition, and their impact on health and disease differs by sex (Bredella, 2017, Brown *et al.*, 2018, Rask-Andersen *et al.*, 2023, Schorr *et al.*, 2018). Although BMI is the most commonly used measure of adiposity, it does not account for fat distribution or sex-specific differences in body composition of fat distribution (Rask-Andersen *et al.*, 2023). Women typically have higher body fat percentages than men of the same BMI, while men have greater lean mass (Bredella, 2017). Men are more prone to accumulate adipose tissue around the trunk and abdomen, whereas women tend to store fat around the hips and thighs (Bredella, 2017, Rask-Andersen *et al.*, 2023). Genome-wide association studies on adiposity suggest genetic variants may also have sex-specific effects, which may be mediated by sex-related factors such as genetics and sex hormones (Bredella, 2017). While both men and women are susceptible to obesity, the patterns of fat

distribution, incidence, and health consequences may differ between the sexes (Lemieux *et al.*, 1993, Power and Schulkin, 2008).

Rask-Andersen *et al.* (2023) assessed adiposity-related traits beyond BMI and their associations with cancer risk in men and women using data from the UK Biobank. They reported adiposity-related traits were associated with cancer risk differently in men and women, and were particularly notable for cancers of the oesophagus, colorectum and liver cancers. For colorectal cancer, they noted visceral to subcutaneous adiposity ratio (indicating higher visceral adiposity) was associated with increased risk in men but not women. For cancers of the liver and oesophagus, they noted sex-interactions were significant for various adiposity measures (including BMI), and traits were more strongly associated with increased risk of cancer in men but not in women. Overall, their findings suggested that the effects of adiposity—particularly visceral adiposity—on cancer risk may differ by sex, contributing to higher risk for certain cancers among men and supporting the need for sex-stratified analyses (Rask-Andersen *et al.*, 2023).

**Table 2.1 Body mass index (BMI) for adults in the UK (England) by age and sex.**

| | 35-44 | 45-54 | 55-64 | 65-74 | 75+ | Total |
|---|---|---|---|---|---|---|
| Men | | | | | | |
| Mean BMI | 27.8 | 28.8 | 28.6 | 28.5 | 27.5 | 27.5 |
| SE | 0.30 | 0.26 | 0.24 | 0.21 | 0.20 | 0.12 |
| % Normal BMI | 30 | 18 | 21 | 24 | 27 | 31 |
| % Overweight | 44 | 52 | 47 | 45 | 49 | 43 |
| % Obese | 26 | 30 | 33 | 31 | 24 | 25 |
| % Overweight & obese | 70 | 82 | 79 | 76 | 73 | 69 |
| Women | | | | | | |
| Mean BMI | 27.7 | 28.0 | 28.2 | 28.2 | 27.6 | 27.2 |
| SE | 0.34 | 0.30 | 0.30 | 0.27 | 0.29 | 0.14 |
| % Normal BMI | 38 | 35 | 36 | 29 | 35 | 41 |
| % Overweight | 34 | 35 | 33 | 39 | 37 | 32 |
| % Obese | 28 | 30 | 31 | 32 | 27 | 26 |
| % Overweight & obese | 62 | 65 | 64 | 71 | 65 | 59 |

The Health Survey for England is a series of surveys commissioned by NHS Digital and carried out by NatCen Social Research and University College London. The surveys are representative of adults and children in England and are used to monitor the nation's health and health-related behaviours.
BMI status: Not overweight nor obese ('normal') <25kg/m2; Overweight ≥25 to <30kg/m2; Obese ≥30kg/m2. BMI calculated from self-reported height and weight.
SE, standard error of the mean
Table data sourced from The Health Survey for England, 2021 (NHS Digital, 2022a).

## 2.3.3 Genetic Factors

The contributions from X and Y chromosomes predominantly make up the genetic sex differences in men and women, which persist throughout life (Mauvais-Jarvis *et al.*, 2020). The molecular and cellular characteristics of all male and female cells are driven by the differences in composition and regulation of the X and Y chromosomes (Mauvais-Jarvis *et al.*, 2020). The basis of sex differences in men and women stem from the X and Y chromosomes, which are reflected physiologically by sex hormones and anatomically by the gonads, external genitalia and intrinsic reproductive organs (Clayton and Tannenbaum, 2016, Haupt *et al.*, 2021). Efforts to integrate X and Y chromosome sequencing into GWAS have only recently begun, presenting new opportunities for uncovering sex-linked genetic risk factors in cancer.

## X Chromosomes

The X chromosome is around three times larger than the Y chromosome and has over 800 protein coding genes (Miga *et al.*, 2020, National Human Genome Research Institute, 2024). It is hypothesized that, to an extent, women may have conferred protection against cancer because they carry two X chromosomes (Dunford *et al.*, 2017, Haupt *et al.*, 2021). As female cells inherit two copies of the X chromosome, one copy undergoes random inactivation during embryogenesis to prevent sex differences in X chromosome gene dosage 2020 (Mauvais-Jarvis *et al.*, 2020). The non-recombining portion of the X chromosome contains tumor suppressor genes, some of which may escape inactivation, leading to biallelic expression (Dunford *et al.*, 2017). The mechanisms underlying this inactivation process remain understudied, but biallelic expression of certain X-linked genes has been observed across individuals and cell types (Dunford *et al.*, 2017). These group of genes have been termed EXITS genes, derived from 'escape from X inactivation tumour suppressors' (Dunford *et al.*, 2017). A study in The Cancer Genome Atlas aimed at identifying EXITS genes reported six of 783 genes on the non-recombination portion of the X chromosome (*ATRX, CNKSR2, DDX3X, KDM5C, KDM6A* and *MAGEC3*) harboured loss-of-function mutations more frequently in men than in women and thus hypothesized male-biased mutations may contribute to increased cancer risk in men (Dunford *et al.*, 2017). Notably, the *KDM6A* gene, also known as *UTX*, is reportedly frequently mutated in bladder cancer patients and suggested to play an important role in its tumorigenesis, for which there is a disproportionately high incidence in men (Chaudhary *et al.*, 2025, Ler *et al.*, 2017).

## Y chromosome

The Y chromosome is about one third in size compared to the X chromosome and has around 100 protein coding genes (Dirican and Nelson, 2024, Rhie *et al.*, 2023). It is understood to have evolved alongside the X chromosome millions of years ago, albeit a significant proportion of its sequence degraded with gradual loss of recombination potential, retaining only three to

five percent of its original length (Dirican and Nelson, 2024, Haupt *et al.*, 2021). The Y chromosome was only recently sequenced telomere-to-telomere, and until then its sequence had 50% of its bases represented by gaps in the human reference genome (Li and Durbin, 2024, Rhie *et al.*, 2023). This was due to the complex architecture of the chromosome, as its non-recombination portion comprises of extensive regions of base repeats, leading to complex genetic structures of tandem repeats, palindromes, and segmental duplication (Li and Durbin, 2024, Rhie *et al.*, 2023). Research on the direct association between the Y chromosome and cancer risk is still emerging (Rojas *et al.*, 2024).

The mosaic loss of Y chromosome (mLOY) is a frequently occurring somatic aneuploidy in normal haematopoietic cells in men and understood to be a function of ageing (Bruhn-Olszewska *et al.*, 2025). It gained importance in the last decade as it was first associated with increased risk of non-haematological malignancies (Bruhn-Olszewska *et al.*, 2025, Forsberg *et al.*, 2014). Since 2014, over 160 studies have reported the association of mLOY and disease risk, including cancers of the head and neck, lung, and bladder (Bruhn-Olszewska *et al.*, 2025). It is hypothesized the male predominance in cancer risk may be attributed to mLOY directly, or through ageing and environmental risk factors such as smoking and exposure to carcinogens that promote mLOY (Bruhn-Olszewska *et al.*, 2025). Genes located on the non-recombination male-specific region of the Y chromosome such as *DDX3Y*, *EIF1AY*, *KDM5D*, *RPS4Y1*, *UTY* (also known as *KDM6C*), *ZFY* and *TMSB4Y* have been suggested to be tumour suppressor genes (Bruhn-Olszewska *et al.*, 2025). Their loss may confer a proliferative advantage to cells, contributing to tumour formation (Bruhn-Olszewska *et al.*, 2025). Additionally, although less frequently than mLOY, men also experience partial or complete mosaic loss of the X chromosome in blood cells, which may exacerbate cancer risk (Bruhn-Olszewska *et al.*, 2025).

### Autosomal Differences

Before advances in genetic epidemiology allowed for sequencing of the X and Y chromosomes, research on sex differences primarily focused on autosomal variation. Sex differences in autosomal somatic cancer genes are thought to result from a combination of sex hormone-related transcriptional regulation, transcription factor influence, and epigenetic regulation (Haupt *et al.*, 2021, Lopes-Ramos *et al.*, 2020). Notably, sex differences in gene expression have been observed across 44 tissue types in healthy individuals (Haupt *et al.*, 2021). A study analysing 28 cancer subtypes as part of the Pan-Cancer Analysis of Whole Genomes (PCAWG) project identified sex differences in coding and non-coding mutations, as well as genome-wide phenomena such as mutational signature activity (Li *et al.*, 2020).

With growing efforts to meta-analyse cancer genetic datasets and perform genome-wide association studies (GWAS), there has been an increase in sex-specific GWAS for cancers with significant sex disparities (Koutros *et al.*, 2023, Laskar *et al.*, 2019, Purdue *et al.*, 2024). These studies have identified genetic associations that exhibit sex differences, suggesting potential disparities in heritability and/or genetic correlation between sexes.

## 2.3.4 Sex Hormones

As highlighted in the introduction of the thesis (Chapter 1, Section 1.2), considerable research has examined the influence of sex hormones on cancer incidence in men and women. Nearly every cancer site with a male or female predominance in its incidence has been hypothesized to be driven by sex hormones to some extent. Sex steroid hormones are the primary drivers of the development and homeostasis of secondary sexual characteristics, and are understood to mediate carcinogenesis-related events including cell proliferation, differentiation, metabolism, inflammation, and apoptosis (Clocchiatti *et al.*, 2016, Li *et al.*, 2019, Lopes-Ramos *et al.*, 2020, Mauvais-Jarvis *et al.*, 2020, Petrick and Cook, 2020, Roy *et al.*, 2025, Zhao

*et al.*, 2024). However, direct comparisons of the effects of sex-related biological factors on sex differences in cancer incidence remain difficult to isolate, as illustrated with the example of breast cancer earlier (Chapter 1, Section 1.2). Here, some studies that have evaluated the role of sex hormones and risk of oesophageal adenocarcinoma, a cancer with a notable sex disparity (reportedly over five-fold greater incidence in men compared to women) is considered.

The persistent male predominance of oesophageal adenocarcinoma (OAC) has long generated hypotheses surrounding the role of sex hormones towards its aetiology, despite it not being perceived as a hormone-dependent cancer (Abnet *et al.*, 2018, O'Doherty *et al.*, 2012, Petrick and Cook, 2020). It was initially hypothesized testosterone may increase risk of OAC, and oestrogen may be protective (Abnet *et al.*, 2018, Cook *et al.*, 2015, Cooper *et al.*, 2009, Green *et al.*, 2012). A study among 100,000 men receiving oestrogen treatment for prostate cancer reported no evidence of any protective effect on OAC risk (Lagergren and Nyrén, 1998). This hypothesis that female sex hormones may be protective was further explored in women receiving hormone replace therapy, but also showed no evidence in support of this theory (Lindblad *et al.*, 2006). Contrastingly, a prospective analysis in the Million Women Study reported risk of oesophageal cancer was higher in postmenopausal women compared to pre- or peri-menopausal women (RR 1.46; 95% CI 1.07-2.00, $p<0.01$) (Green *et al.*, 2012). Furthermore, upon assessment of reproductive factors (childbearing, including women's age at first birth, their number of children, breastfeeding history, and parity), they reported risk was higher in nulliparous women compared to parous women (RR 1.31; 1.11-1.55, $p=0.002$) (Green *et al.*, 2012).

Following this, results from one study suggested that higher levels of androgens may be associated with Barrett's oesophagus (BE) in men, a well-established precursor condition for OAC risk. They reported relative risks for highest quartiles of free testosterone (OR 5.36; 95%

CI 2.21-13.03, p=.0002) and free dihydrotestosterone (OR 4.25: 95% CI 1.87-9.66, p=0.001) were associated with BE risk (Cook *et al.*, 2015). Contrastingly, two prospective analyses explored the role of sex hormones and reported inverse associations with sex hormones and OAC risk in men (Petrick *et al.*, 2019, Xie *et al.*, 2020). Xie et al., (2020) used data from the Janus Serum Bank Cohort (Norway), and reported decreased risk of OAC comparing the highest to lowest quartiles of testosterone (OR 0.44; 95% CI 0.22-0.88), testosterone to oestradiol ratio (OR 0.37; 95% CI 0.19-0.72), and luteinizing hormone (OR 0.50, 95% CI 0.30-0.98) in men. Petrick et al., (2019) reported higher concentrations of dehydroepiandrosterone were associated with 38% reduced risk of oesophageal adenocarcinoma and gastric cardia adenocarcinoma (combined as an endpoint) in men. The authors also reported higher oestradiol levels were associated with 34% reduced risk of cancers in men. Differing in their results from the analyses mentioned above, a recent prospective analysis in the UK Biobank of prediagnostic sex hormones in men and women reported no associations with risk of OAC (Mcmenamin *et al.*, 2021).

Taken together, oesophageal adenocarcinoma highlights the complexities of understanding how sex hormones may contribute to its pronounced sex differences, a cancer site for which sex hormone-related hypotheses are particularly compelling.

## 2.4 Sex Differences in Incidence by Cancer Site

**Table 2.2 Cancer incidence age-standardized rates per 100,000 person-years, for men and women, aged 35 years and above, diagnosed in UK between 2013-2017.**

| Cancer sites | Men | Women |
|---|---|---|
| All sites[a] | 760.3 | 674.4 |
| Oral cavity | 29.2 | 12.3 |
| Oesophagus | 25.4 | 9.0 |
| Stomach | 16.4 | 7.2 |
| Colorectum | 93.5 | 65.0 |
| Colon | 55.4 | 42.4 |
| Rectum and anus | 38.0 | 22.5 |
| Liver | 15.9 | 6.5 |
| Gallbladder | 3.6 | 4.1 |
| Pancreas | 20.1 | 15.6 |
| Lung | 94.9 | 74.9 |
| Melanoma | 33.9 | 32.5 |
| Thyroid | 4.5 | 11.1 |
| Kidney | 30.3 | 15.3 |
| Bladder | 30.6 | 9.7 |
| Hodgkin lymphoma | 4.2 | 2.4 |
| Non-Hodgkin lymphoma | 30.3 | 21.4 |
| Multiple myeloma | 15.3 | 10.2 |
| Leukaemia | 22.9 | 13.0 |

An age-standardized incidence rate (ASIR) is a weighted mean of the age-specific rates; the weighting is based on the population distribution of a standard population. The most frequently used standard population is the World Standard Population, and the ASIR is expressed per 100,000 person-years (Ervik et al., 2024). Cancer incidence data include England, Northern Ireland, Scotland and Wales.

[a]All sites excluding non-melanoma skin cancer.

As part of the literature review, ASIRs per 100,000 person-years from this table are reported in the text for the cancer sites reviewed. For selected cancer subsites where GLOBOCAN data were unavailable, ASIRs were obtained from additional sources.

Source: Global Cancer Observatory (Ervik et al., 2024)

### 2.4.1 Oral cavity (C00-C14)

The term ‘cancers of the oral cavity’ has been used throughout this thesis to describe malignant neoplasms of lip, oral cavity and pharynx (Appendix Table 3C). Some studies also refer to this group of cancers as head and neck cancers or oropharyngeal cancers (Chaturvedi *et al.*, 2022, Gaeta *et al.*, 2023, Gillison *et al.*, 2015). While each cancer subsite exhibits heterogeneity in tumour types, grouping into ‘oral cavity’ is done based on three main shared risk factors—smoking, alcohol, and Human Papillomavirus (hr-HPV, especially HPV16) infection (Chaturvedi *et al.*, 2022, Gaeta *et al.*, 2023, Gillison *et al.*, 2015). Additional established risk factors include occupations, ionising radiation exposure, and poor oral health (Brown *et al.*, 2018, Chaturvedi *et al.*, 2022) Altogether, these known risk factors are suggested to contribute between 45% to 90% of the oral cavity incidence in the UK (Brown *et al.*, 2018). Between 2013 and 2017, ASIRs per 100,000 person-years were 29.2 in men and 12.3 in women (Table 2.2).

The risk of cancers of the oral cavity is two- to three- times greater in men compared to women and is hypothesized to be due to sex biased distribution of key risk factors (Brown *et al.*, 2018, Gillison *et al.*, 2015, Jackson *et al.*, 2022, Shelton *et al.*, 2024). While smoking rates are declining in men, the incidence of oral cavity cancers has continued to rise particularly for young white men (<60 years), further widening the gap between the sexes (Chaturvedi *et al.*, 2022, Shelton *et al.*, 2024). It is hypothesized to be due to increased oral HPV exposure related to changes in oral sexual behaviours among men (Chaturvedi *et al.*, 2022, Gillison *et al.*, 2015). The incidence of oral cancers has increased for women too, however, not as high among men, and may be attributable to rises in smoking rates and cohort differences in tobacco exposure, as women adopted smoking more gradually than men (Brown *et al.*, 2018, Gillison *et al.*, 2015).

Given its relevance to several cancer sites discussed in this section, the study by Jackson *et al.* (2022) is referred to repeatedly for several cancer sites in subsequent sections. Jackson et al. examined sex differences in cancer incidence across 21 cancer sites in the NIH-AARP Diet and Health Study (with 294,100 participants). They first estimated age-adjusted male-to-female incidence rate ratios (MF IRRs) using Poisson regression models. They then used Cox proportional hazards regression models, with length of follow-up as the time scale, to estimate multivariable-adjusted male-to-female hazard ratios (HRs). Risk factors were initially selected for inclusion in full regression models based on their univariate association with each cancer site, using a threshold of $p<0.10$, and were subsequently removed using backward stepwise selection to create final parsimonious models. Covariates were retained in the final models where $p<0.05$ according to the Wald test. Because the age-adjusted and multivariable-adjusted estimates were derived from different modelling frameworks, the estimates from Jackson et al., are described in this review chapter.

Jackson *et al.* (2022) assessed sex differences in oral cavity cancer (n=407 cases; 33.9% in women). They reported an age-adjusted male-to-female incidence rate ratio (MF IRR) of 1.50 (95% confidence interval (CI) 1.22-1.85) and a multivariable-adjusted hazard ratio (HR) of 1.38 (95% CI 0.99-1.91) after adjustment for age, non-linear age effects, marital status, height, non-linear height effects, fibre consumption, alcohol use, and cigarette smoking.

## 2.4.2 Oesophagus [C15]

Oesophageal cancers have the highest incidence rate in the UK compared to any other European country, and their prevalence continues to rise (Gatenby *et al.*, 2011, Lepage *et al.*, 2008). Between 2013 and 2017, ASIRs per 100,000 person-years for oesophageal cancers were 25.4 in men and 9.0 in women (Table 2.2). Over 95% of all oesophageal cancer cases are of two histological subtypes—oesophageal squamous cell carcinoma (OSCC) and oesophageal adenocarcinoma (OAC) (Smyth *et al.*, 2017). While sharing an anatomical site,

OAC and OSCC are distinct diseases with differing cells of origin, risk factors, pathology, epidemiology and incidence patterns (Smyth *et al.*, 2017).

The incidence of OAC is notably higher and increasing in countries with high development indexes, such as the UK (Smyth *et al.*, 2017). It is also characterised by one of the largest sex disparities in cancer incidence, with men having an approximately sevenfold greater risk than women (Smyth *et al.*, 2017). Comparatively, the incidence rate of OSCC is nearly four times greater in men than women (Arnold *et al.*, 2015). In 2012, the ASIRs per 100,000 person-years were 3.4 in men and 0.6 in women for OAC, and 3.7 in men and 1.3 in women for OSCC (Arnold *et al.*, 2015).

Risk factors for OAC include gastro-oesophageal acid or bile reflux, gastro-oesophageal reflux disease (GORD), Barrett's oesophagus (a complication of chronic GORD), and excess adiposity (specifically excess central and visceral adiposity) (Abnet *et al.*, 2018, Rask-Andersen *et al.*, 2023). Key factors for OSCC include tobacco, alcohol consumption and achalasia, and additional risk factors with suggestive evidence based on substantial data include low fruit and vegetable intake, poor oral health, consumption of high-temperature beverages, and human papillomavirus (HPV) infections (Abnet *et al.*, 2018, Smyth *et al.*, 2017).

The male predominance of OAC is less well understood but thoroughly reported, whereas the observed sex differences in differences in OSCC incidence are often attributed to differences in smoking and alcohol use (Arnold *et al.*, 2020, Arnold *et al.*, 2017, Arnold *et al.*, 2015). Jackson *et al.* (2022) investigated sex differences in OAC and OSCC risk by adjusting for several covariates and reported sex differences in OSCC incidence attenuated (HR: 1.46, 95% CI: 0.98–2.18), but the male predominance of OAC for remained after adjusting for several risk factors (HR: 10.80, 95% CI: 7.33–15.90). It has been hypothesised that the greater accumulation of abdominal adiposity, which is more prevalent in men than in women, leads

to increased levels of gastro-oesophageal reflux disease (GORD) and consequently a higher risk of oesophageal adenocarcinoma (OAC) in men (Cook *et al.*, 2005, Coupland *et al.*, 2012). Barrett's oesophagus, a precursor lesion on the causal pathway OAC, is also strongly associated with abdominal adiposity and is more common in men; this is therefore thought to further contribute to the excess risk of OAC observed in men (Cook *et al.*, 2005, Coupland *et al.*, 2012).

Jackson *et al.* (2022) assessed sex differences in OAC (n=469 cases; 6.0% in women) and OSCC (n=147 cases; 31.3% in women). For OSCC, they reported an age-adjusted MF IRR of 1.67 (95% CI 1.18-2.37) and a multivariable-adjusted HR of 1.46 (95% CI 0.98-2.18) after adjustment for age, non-linear age effects, BMI, race, self-reported health status, caloric intake, alcohol use, and cigarette smoking. For OAC, they reported an age-adjusted MF IRR of 12.19 (95% CI 8.32-17.86) and a multivariable-adjusted HR of 10.80 (95% CI 7.33-15.90) after adjustment for age, non-linear age effects, BMI, race, antacid use, first-degree relative with cancer, and cigarette smoking.

### 2.4.3 Stomach [C16]

Stomach (gastric) cancer is the fifth most commonly diagnosed cancer and the fourth leading cause of cancer-related mortality worldwide (Ilic and Ilic, 2022). Its incidence shows a marked sex disparity, with rates approximately twice as high in men as in women (Ilic and Ilic, 2022).

Between 2013 and 2017, ASIRs per 100,000 person-years were 16.4 for men and 7.2 for women (Table 2.2). Stomach cancers are classified into two histological subtypes—gastric cardia (arising from the upper stomach) and gastric non-cardia (arising from other parts of the stomach)—and are distinct diseases by aetiology and epidemiology (Ilic and Ilic, 2022).

The well-established risk factors for stomach cancers are *Helicobacter pylor*i (*H. pylori*) infection (specific to gastric non-cardia cancer with up to 90% of global cancer cases

attributable), tobacco smoking, and ionizing radiation (Song *et al.*, 2022, Ye *et al.*, 2018). Risk factors with evidence based on substantial data include alcohol consumption, high-sodium diets, exposures (e.g., farming, mining, cement work), family history, and partial gastrectomy (for both histological subtypes), and gastroesophageal reflux disease and obesity (for gastric cardia cancer) (Song *et al.*, 2022, Ye *et al.*, 2018). Between 1998 and 2007 in the UK, age-standardized rates of stomach cancer declined in both men and women, with a more pronounced decrease observed in men (Coupland *et al.*, 2012). For gastric non-cardia cancers, ASIRs per 100,000 person-years decreased from 4.0 to 2.4 in men and 1.9 to 1.7 in women (Coupland *et al.*, 2012). The decline in gastric non-cardia cancer may be associated with the decreasing prevalence of *H. pylori* infections (Coupland *et al.*, 2012). For gastric cardia cancers, ASIRs per 100,000 person-years fell from 5.7 to 4.2 in men and 1.4 to 1.1 in women (Coupland *et al.*, 2012). Sex differences in gastric cardia cancers are marked (compared to gastric non-cardia), with up to four-fold greater incidence in men compared to women (Coupland *et al.*, 2012). Well-established risk factors for gastric cardia cancers include alcohol, tobacco smoking, abdominal adiposity, type 2 diabetes, partial gastrectomy and dietary factors (Giovannucci *et al.*, 2010, Liu *et al.*, 2025, Ye *et al.*, 2018). Additionally, there is some overlap in risk factors between gastric cardia cancer and oesophageal adenocarcinoma (OAC)—including gastro-oesophageal reflux disease (GORD), Barrett's oesophagus, and abdominal adiposity—and a similarly marked male predominance observed for both cancers (Coupland *et al.*, 2012). It is therefore hypothesised that sex-differences in the prevalence of well-established (and shared risk factors with OAC) may contribute to the pronounced sex disparities of gastric-cardia cancers.

Jackson *et al.* (2022) assessed sex differences in gastric cardia cancer (n=325 cases; 13.5% in women) and gastric non-cardia cancer (n=347 cases; 27.4% in women). For gastric non-cardia cancer, they reported an age-adjusted MF IRR of 2.04 (95% CI 1.61-2.58) and a multivariable-adjusted HR of 2.21 (95% CI 1.72–2.84) after adjustment for age, non-linear age

effects, race, ibuprofen use, physical activity, alcohol use, cigarette smoking, and coffee/tea consumption. For gastric-cardia cancer, they reported an age-adjusted MF IRR of 4.93 (95% CI 3.59-6.77) and a multivariable-adjusted HR of 3.49 (95% CI 2.26-5.37) after adjustment for age, non-linear age effects, education, antacid use, diabetes, height, non-linear height effects, coffee/tea consumption, vegetable consumption, and cigarette smoking.

### 2.4.4 Colorectum [C18-C20]

In the UK, colorectal cancer ranks as the third most common cancer overall, following lung and prostate cancer in men, and breast and lung cancer in women (Cancer Research UK, 2025, Murphy *et al.*, 2011, White *et al.*, 2018). Between 2013 and 2017, ASIRs per 100,000 person-years were 93.5 in men and 65.0 in women (Table 2.2). Sex differences in colorectal cancer incidence vary by both anatomical subsite and age (Araghi *et al.*, 2019, Cancer Research UK, 2025, Murphy *et al.*, 2011, White *et al.*, 2018, Wong *et al.*, 2021). Among individuals diagnosed before 50 years of age, overall colorectal cancer incidence is similar in men and women; however, when restricted to colon cancer, incidence is higher in women than in men (Araghi *et al.*, 2019, White *et al.*, 2018, Wong *et al.*, 2021). In contrast, at older ages (70–74 years), colorectal cancer incidence is markedly higher in men, reaching nearly twice that observed in women (Araghi *et al.*, 2019, White *et al.*, 2018, Wong *et al.*, 2021).

Well-established risk factors for colorectal cancer include overweight and obesity, consumption of red and processed meat, alcohol intake, tobacco smoking, physical inactivity, family history, and genetic predisposition due to inherited syndromes (such as familial adenomatous polyposis) (Araghi *et al.*, 2019, Huxley *et al.*, 2009, Keum *et al.*, 2018). Additionally, based on substantial evidence, inflammatory bowel diseases, Type 2 diabetes mellitus, low dietary fibre intake, high plasma insulin-like growth factor 1 levels are probable causal contributors to colorectal cancer risk (Araghi *et al.*, 2019, Giovannucci *et al.*, 2010, Huxley *et al.*, 2009, Keum *et al.*, 2018). Epidemiological studies have found that associations

between established risk factors and cancer risk vary by anatomical site (e.g., distal colon, proximal colon, rectum), suggesting heterogeneity in aetiology across subsites (Demb *et al.*, 2019, Keum *et al.*, 2018, Murphy *et al.*, 2011). The sex differences in colorectal cancer risk have been hypothesized to be contributed by the sex differences in prevalence of modifiable risk factors.

Jackson *et al*. (2022) assessed sex differences in colon (n=4091 cases; 37.3% in women) and rectal cancer (n=1,377 cases; 32.2% in women). For colon cancer, they reported and age-adjusted MF IRR of 1.29 (95% CI 1.21– 1.37) and a multivariable-adjusted HR of 0.99 (0.90–1.09) after adjustment for age, non-linear age effects, BMI, diabetes, height, non-linear height, ibuprofen use, self- reported health status, fibre consumption, oil consumption, physical activity, red meat consumption, white potato consumption, alcohol use and cigarette smoking. For rectal cancer, they reported an age-adjusted MF IRR of 1.61 (95% CI 1.44-1.80) and a multivariable-adjusted HR of 1.36 (95% CI 1.20-1.54) after adjustment for age, non-linear age effects, education, antacid use, diabetes, ibuprofen use, red meat consumption, alcohol use, and cigarette smoking.

### 2.4.5 Anus [C21]

Although relatively less common, anal cancer incidence has been rising in several countries, including the UK (Wilkinson *et al.*, 2014). Between 1990 and 2010, incidence rose by 69% in men and 126% in women, with women consistently having higher incidence overall (Wilkinson *et al.*, 2014). From 2017 to 2019, age-standardized incidence rates per 100,000 people in the UK were 1.6 for men and 3.1 for women (Cancer Research UK, 2025). Consistent with these trends, cross sectional analyses of national cancer registry data indicate that anal cancer incidence is approximately 19–32% higher in women than in men (Cook *et al.*, 2009, Dong *et al.*, 2020).

Risk factors for anal cancer mainly include HPV infection and human immunodeficiency virus (HIV) infection, and cofactors related to infection exposure—sexually transmitted infections through anal sexual intercourse and high number of lifetime sexual partners (Grulich *et al.*, 2017, Wilkinson *et al.*, 2014). Additionally, smoking, immunosuppression, and genetic risk factors are also contributory (Grulich *et al.*, 2017). Sex differences in prevalence and exposure to HIV and HPV infections are suggested to partially contribute (Grulich *et al.*, 2017).

A cross-sectional study using US SEER data assessing sex differences in cancer incidence of 49 sites across ethnicities reported women had higher risk of anal cancer across most racial and ethnic groups, except amongst Black individuals where men had higher risk than women (Tosakoon *et al.*, 2024). The authors inferred that this pattern may be driven by HIV prevalence, which is higher among Black and Latino men with anal cancer than among White men, and that HIV-associated anal cancer contributes substantially to anal cancer incidence in men but not in women, with approximately 35% of men with anal cancer being HIV-positive compared with around 2% of women (Tosakoon *et al.*, 2024, Zhang *et al.*, 2022).

Jackson *et al*. (2022) assessed sex differences in anal cancer (n=148 cases; 44.6% in women). They reported an age-adjusted MF IRR of 0.97 (95% CI 0.70-1.34) and a multivariable-adjusted HR of 1.17 (95% CI 0.82-1.69) after adjustment for age, non-linear age effects, and marital status.

### 2.4.6 Liver [C22]

Liver cancer ranks as the sixth most common cancer worldwide and the third leading cause of cancer-related deaths (Ervik *et al.*, 2024, International Agency for Research on Cancer, 2025a). Compared to global patterns, liver cancer is relatively less common in the UK as over 80% of cases occur in less-developed countries (Burton *et al.*, 2021, Wong *et al.*, 2021). Nonetheless, sex differences in incidence are notable; with men having a two- to four-fold

higher risk compared to women (Burton *et al.*, 2021). Between 2013 and 2017, ASIRs per 100,000 person-years were 15.9 in men and 6.5 in women (Table 2.2).

Well-established risk factors for liver cancer primarily include alcohol consumption and alcohol-related diseases, and chronic hepatitis B and C viral infections (Burton *et al.*, 2021, Wong *et al.*, 2017). The observed sex differences in incidence are also hypothesized to be contributed by these two risk factors (Bamia *et al.*, 2018, Wong *et al.*, 2017). Additional important risk factors include tobacco smoking, obesity, gallbladder disease, diabetes, non-alcoholic fatty liver disease, and exposure to aflatoxin B (Bamia *et al.*, 2018, Burton *et al.*, 2021, Lonardo *et al.*, 2019, Wong *et al.*, 2017). Studies investigating sex differences in liver cancer additionally hypothesize sex hormones—particularly testosterone and sex hormone-binding globulin (SHBG)—may potentially contribute to the male predominance of hepatocellular carcinoma, the most common liver histological subtype (>80% of cases), although this warrants further research (Bamia *et al.*, 2018, Watling *et al.*, 2025, Wong *et al.*, 2017).

Jackson *et al.* (2022) assessed sex differences in liver cancer (n=539 cases; 24.7% in women). They reported an age-adjusted MF IRR of 2.38 (95% CI 1.96-2.89) and a multivariable-adjusted HR of 2.60 (95% CI 2.06-3.28) after adjustment for age, non-linear age effects, BMI, marital status, race, aspirin use, diabetes, self-reported health status, red meat consumption, white meat consumption, alcohol use, and cigarette smoking.

### 2.4.7 Gallbladder [C23]

Cancer of the gallbladder is the most common cancer of the biliary tract, and one of the few cancer sites with a higher incidence in women compared to men (Bamia *et al.*, 2018, Shaffer and Hundal, 2014, Torre *et al.*, 2018). Gallbladder cancer incidence varies regionally, being the highest in South America, northern India, and East Asia, and relatively rare in the UK (Roa *et al.*, 2022, Torre *et al.*, 2018). Between 2013 and 2017, ASIRs per 100,000 person-years were

3.6 in men and 4.1 in women in the UK (Table 2.2). Overall, incidence is thought to be approximately twice as high in women, although substantial regional variability exists in sex-specific incidence, with ratios ranging from 1:1 in the Far East to 4:1 in Spain (Jackson *et al.*, 2020, Torre *et al.*, 2018).

The primary risk factors for cancer are gallstones and gallbladder disease in (Jackson *et al.*, 2020, Torre *et al.*, 2018). Additional risk factors include obesity, dietary risk factors, and hormonal factors, that are also understood to strongly influence the development of gallstones (Bamia *et al.*, 2018). Sex differences in cancer incidence have been linked to hormonal factors (multiple pregnancies and exogenous oestrogen use) and their associations with gallstones and gallbladder disease in (Jackson *et al.*, 2020, Torre *et al.*, 2018).

Jackson *et al.* (2022) assessed sex differences in gallbladder cancer (n=110 cases; 54.6% in women). They reported an age-adjusted MF IRR of 0.65 (0.44– 0.94) and a multivariable-adjusted HR of 0.33 (95% CI 0.18– 0.58) after adjustment for age, non-linear age effects, race, BMI, height, non-linear height effects and red meat consumption.

### 2.4.8 Pancreas [C25]

Pancreatic cancer is the eleventh most common cancer and the seventh leading cause of cancer-related deaths worldwide (Rawla *et al.*, 2019). Between 2013 and 2017, ASIRs per 100,000 person-years were 20.1 for men and 15.6 for women in the UK (Table 2.2). Established risk factors include smoking, alcohol consumption, obesity, type 2 diabetes, family history, genetic predisposition and occupational exposures to substances like cadmium and arsenic (Rawla *et al.*, 2019). Epidemiological studies assessing national-level data report incidence for men may be on average 30% higher incidence compared to women (Cook *et al.*, 2009, Dong *et al.*, 2020). The excess risk is hypothesized to be associated higher rates of smoking and alcohol consumption (Rawla *et al.*, 2019). Unlike other cancer sites, sex differences in pancreatic cancer incidence have been inconsistently reported and the excess risk in men is

comparatively modest (Cook *et al.*, 2009, Dong *et al.*, 2020). Jackson *et al.* (2022)quantified sex differences in pancreatic cancer after accounting risk factors observed a risk ratio attenuated to the null (HR: 1.06, 95% CI: 0.91–1.23).

Jackson *et al.* (2022) assessed sex differences in pancreatic cancer (n=1,657 cases; 38.2% in women). They reported an age-adjusted MF IRR of 1.26 (95% CI 1.14-1.39) and a multivariable-adjusted HR of 1.06 (95% CI 0.91-1.23) after adjustment for age, non-linear age effects, education, diabetes, self-reported health status, white potato consumption, and cigarette smoking.

### 2.4.9 Lung [C34]

Lung cancer is the most commonly diagnosed cancer in men and the main neoplastic cause of death in both sexes (Malhotra *et al.*, 2018). It is estimated to contribute to 18% of all cancer deaths (Sung *et al.*, 2021). Between 2013 and 2017, ASIRs per 100,000 person-years were 94.9 for men and 74.9 for women in the UK (Table 2.2). Based on cancer incidence from 2015 in the UK, the population attributable fractions of all risk factors combined for lung cancer incidence was 78.9%, of which 70.0% cases were attributable to smoking (Brown *et al.*, 2018). Lung cancer is classified by two main histological subtypes: small cell lung carcinoma (SCLC), which comprises for approximately 5-15% of all cases, and non-small cell lung carcinoma (NSCLC), which comprises 75-85% of cases (Inamura, 2017, Malhotra *et al.*, 2018). NSCLC includes a group of malignant non-small cell epithelial tumours that are further categorized into three main subtypes; adenocarcinoma (40% of NSCLC cases), squamous cell carcinoma (25-30%), and large cell carcinoma (10% of NSCLC cases) (Inamura, 2017, Malhotra *et al.*, 2018). Squamous cell carcinomas are mostly located centrally in the larger bronchi of the lung and also linked more strongly with smoking than other subtypes of NSCLC (Jeon *et al.*, 2025, PDQ Adult Treatment Editorial Board, 2025). Lung adenocarcinomas are histologically heterogenous and comprise of mixtures of subtypes including acinar, papillary,

bronchioloalveolar, and solid adenocarcinoma with mucin formation, and their criteria for diagnosis has had heterogeneity in the past (PDQ Adult Treatment Editorial Board, 2025). Large cell carcinomas are also a histologically heterogenous group and defined as a tumour lacking morphologic or immunophenotypic differentiation characteristics of squamous cell carcinoma, adenocarcinoma, or neuroendocrine carcinoma (Sholl, 2016). Lung cancer incidence trends vary considerably by histological subtypes and are associated with various risk factors, predominantly including the effects of smoking on lung cancer risk (Jeon *et al.*, 2025). Additionally, risk factors include changes in cigarette composition, diagnostic technological advances, variations in environmental and occupational exposures associated with subtypes, and lung cancer screening methods (Jeon *et al.*, 2025, Pinsky *et al.*, 2013).

All lung cancer subtypes except adenocarcinoma exhibit higher incidence and mortality rates in men compared to women, and the magnitude of sex differences are largely attributable to the sex differences in smoking prevalence (Malhotra *et al.*, 2018). Since the 1990s, as smoking prevalence has declined in high-income countries, lung adenocarcinoma has become the most common subtype of lung cancer. and additionally, the predominant lung cancer subtype in women, but also never-smokers (Jemal *et al.*, 2018, Jemal *et al.*, 2025, Lopiccolo *et al.*, 2024). This shift has been particularly pronounced in younger women compared to younger men and varies across ethnicities and regional populations (Jemal *et al.*, 2018, Jemal *et al.*, 2025). These contrasting patterns largely reflect sex and cohort differences in tobacco exposure, as women adopted smoking more gradually during the 1950s than men, and the decline in smoking prevalence has been slower and more delayed among women (Brown *et al.*, 2018, Peters *et al.*, 2014, Peto *et al.*, 1976). It is estimated the impact of smoking on lung cancer incidence, especially among women may continue to rise in the coming years (Huang *et al.*, 2022b, Peto *et al.*, 2000, Shelton *et al.*, 2024). Additionally, changes in cigarette composition from nonfiltered cigarettes to filtered and low-tar cigarettes, advances in diagnostic imaging technology, exposure to air pollution and environmental toxins, and

improvements in histological classifications of NSCLC are also considered to contribute to the rise in lung adenocarcinoma incidence (Hill *et al.*, 2023, Jeon *et al.*, 2025, Song *et al.*, 2017).

Occupational carcinogenic agents and environmental carcinogenic agents are significant risk factors for lung cancer incidence after smoking (Malhotra *et al.*, 2018). These include exposure to arsenic and arsenic compounds, asbestos, air pollution levels beryllium, cadmium, chromium compounds, involuntary tobacco smoke, nickel compounds, outdoor air pollution (particularly fine particular matter; $PM_{2.5}$), radon, silica, coal-tar pitch, soot, and additionally occupations involving aluminium production, working with coal, coke production, and working in painting, mining, iron and steel founding, and rubber industries (Malhotra *et al.*, 2018).

Jackson *et al.* (2022) assessed sex differences in lung cancer (n=8,153 cases; 39.6% in women). They reported an age-adjusted MF IRR of 1.18 (95% CI 1.13-1.23) and a multivariable-adjusted HR of 0.99 (95% CI 0.92-1.06) after adjustment for age, non-linear age effects, BMI, education, race, first-degree relative with cancer, height, non-linear height effects, self-reported health status, coffee/tea consumption, fruit consumption, grain consumption, red meat consumption, white meat consumption, alcohol use, cigar/pipe smoking, and cigarette smoking.

Chapter 6 on investigating the female predominance of lung adenocarcinoma in never-smokers describes its epidemiology and reviews its established and emerging risk factors.

### 2.4.10 Malignant melanoma [C43]

Malignant melanoma (hereafter referred to as melanoma) was the fifth most common cancer in the UK in 2017-2019 (Cancer Research UK, 2025). Between 2013 and 2017, ASIRs per 100,000 person-years were 33.9 for men and 32.5 for women (Table 2.2).

Exposure to ultraviolet (UV) radiation is the primary risk factor for melanoma incidence (Brown *et al.*, 2018, Green *et al.*, 2018). Notably, sex differences in melanoma incidence vary across age groups and cancer sites. Women in younger age groups tend to have higher melanoma rates than men, while men surpass women in mid-life to older age groups (Olsen, 2020). The age at which this shift occurs depends on the exposure to UV radiation levels specific to each country. For instance, in Australia, where UV radiation is highest, the switch is noted to occur between ages 45–49, whereas in Denmark, it occurs later, between ages 65–69 (Olsen, 2020).

Anatomic sites of melanoma also vary by sex. In men, higher incidences of melanoma are observed on the head, neck, and upper limbs, particularly at older ages. In contrast, women experience higher rates of melanoma on the trunk and lower limbs, particularly at younger ages (Olsen, 2020). These sex differences are thought to be attributed to sex differences in behavioural factors, such as clothing choices, as well as anatomical differences in the distribution of melanocytes (Olsen, 2020).

Jackson et al. (2022) assessed sex differences in skin cancer, excluding basal and squamous cell carcinoma (n=3102 cases, 26.1% in women. They reported an age-adjusted MF IRR of 2.20 (95% CI 2.03-2.38) and a multivariable-adjusted HR of 1.42 (95% CI 1.26-1.60) after adjustment for age, non-linear age effects, BMI, education, race, height, non-linear height effects, physical activity, vegetable consumption, white meat consumption, alcohol use and cigarette smoking.

### 2.4.11 Kidney [C64-C65]

Kidney cancer ranks as the 14th most common cancer worldwide and the 16th leading cause of cancer-related deaths (Ervik *et al.*, 2024). There is a twofold higher incidence in men compared to women across age groups, time periods, and geographic regions, making it a cancer site with a notable sex disparity (Edgren *et al.*, 2012, Scelo *et al.*, 2018). Between 2013 and 2017, ASIRs per 100,000 person-years were 30.3 for men and 15.3 for women (Table 2.2).

Tobacco smoking and obesity are the primary risk factors associated with kidney cancer, for which distributions in prevalence between sexes have shifted variably over the past 25 years (Scelo *et al.*, 2018, Shelton *et al.*, 2024). Despite this, the sex difference in kidney cancer incidence has remained consistent (Scelo *et al.*, 2018). The stability of the sex difference has been hypothesized to be contributed to by biological (hormones and adiposity) differences between men and women, rather than health-related factors (such as smoking) (Rask-Andersen *et al.*, 2023, Scelo and Larose, 2018).

Jackson *et al.* (2022) assessed sex differences in kidney cancer (n=1,783 cases; 28.0% in women). They reported an age-adjusted MF IRR of 2.00 (95% CI 1.80-2.21) and a multivariable-adjusted HR of 1.57 (95% CI 1.34-1.83) after adjustment for age, non-linear age effects, BMI, education, height, non-linear height effects, self-reported health status, fibre consumption, grain consumption, alcohol use, and cigarette smoking.

### 2.4.12 Bladder [C67]

Bladder cancer incidence exhibits a pronounced sex disparity, with over three-fold higher incidence in men (Burger *et al.*, 2013, Cumberbatch *et al.*, 2018, Jubber *et al.*, 2023, Kogevinas *et al.*, 2018). Between 2013 to 2017, ASIRs per 100,000 person-years were 30.6 in men and 9.7 in women (Table 2.2). The most prominent risk factor for bladder cancer is tobacco smoking, and while a reduction in smoking rates in men has corresponded to a reduction in bladder cancer incidence, the sex gap remains (Kogevinas *et al.*, 2018). Following tobacco smoking, genetic susceptibility, *S. haematobium* infection, occupational carcinogenic exposures, arsenic exposure, family history, and ionizing radiation (Kogevinas *et al.*, 2018). Occupational carcinogen exposures to aromatic amines, polyaromatic hydrocarbons, dyes, substances used in painting, rubber, aluminium, magenta production industries, diesel engine exhausts, and metal working fluids are most consistently associated with increased risk (Kogevinas *et al.*, 2018). The occupations exposed to these carcinogens include being a

painter, machinist, mechanic, dye worker, or worker in the metal industry such as sheet metal work and blacksmithing, work in the textile industry, leather workers, hairdressing, dry cleaning, transport work, waiters, and tobacco workers (Kogevinas *et al.*, 2018). The population attributable fraction from ever smoking are around 66% for men, and 30% for women (Kogevinas *et al.*, 2018). Following smoking, occupational exposures are the second highest contributors to risk, with high risk occupations estimated to be 4-10% for men and slightly lower for women (Kogevinas *et al.*, 2018). Together, it is suggested that higher numbers of men participating in high risk occupations and differences in smoking habits between men and women are important contributors for the sex disparity in incidence (Burger *et al.*, 2013, Cumberbatch *et al.*, 2018, Jubber *et al.*, 2023, Kogevinas *et al.*, 2018).

Jackson *et al.* (2022) assessed sex differences in bladder cancer (n=1,992 cases; 16.3% in women). They reported an age-adjusted MF IRR of 3.96 (95% CI 3.52-4.46) and a multivariable-adjusted HR of 3.33 (95% CI 2.93-3.79) after adjustment for age, non-linear age effects, race, antacid use, diabetes, first-degree relative with cancer, self-reported health status, coffee/tea consumption, fibre consumption, cigar/pipe smoking, and cigarette smoking.

### 2.4.13 Thyroid [C73]

The incidence of cancer of the thyroid has been observed to be consistently higher in women compared to men (Dong *et al.*, 2020, Leclair *et al.*, 2021, Tran and Davies, 2023). In the UK, ASIRs for 100,000 person-years from 2013 and 2017 were 4.5 for men and 11.1 for women (Table 2.2). The incidence of thyroid cancer has continued to increase steadily globally, with a more marked increase in women (Leclair *et al.*, 2021). This increase in incidence is somewhat attributed to improvements in diagnostic methods, such as ultrasounds and fine needle aspiration biopsies (Leclair *et al.*, 2021).

Several risk factors have been associated with thyroid cancer, including goiter, ionizing radiation exposure, particularly during infancy or early childhood, family history, benign nodules or adenomas, iodine intake (Vecchia *et al.*, 2018). Hormonal and reproductive factors are implicated in contributing to the higher risk in women for thyroid related diseases (goiter) and cancers (Vecchia *et al.*, 2018). Female-specific risk factors include menarche, pregnancy, oral contraceptive use, parity, and menopause, however the strength of evidence surrounding them is variable (Leclair *et al.*, 2021, Schubart *et al.*, 2021, Vecchia *et al.*, 2018).

Jackson *et al.* (2022) assessed sex differences in thyroid cancer (n=520 cases; 57.3% in women). They reported an age-adjusted MF IRR of 0.59 (95% CI 0.49-0.70) and a multivariable-adjusted HR of 0.55 (95% CI 0.46-0.66) after adjustment for age, non-linear age effects, BMI, and white meat consumption.

## 2.4.14 Lymphatic and Haematopoietic Cancers [C81-C96]

### Hodgkin Lymphoma [C81]

Hodgkin lymphoma (HL) is a malignancy of lymphocytes, with its vast majority of cases derived from germinal centre B-lymphocytes and few from T-cell origin (Hjalgrim and Jarrett, 2020). Compared to the cancer sites investigated in this thesis, it is a relatively rare cancer, being the 27$^{th}$ most common cancer diagnosed and higher incidence in affluent Western nations compared to Asian nations (Hjalgrim and Jarrett, 2020). Age-specific incidence trends from western populations report low incidence rates of HL in childhood, marked incidence in adolescents and young adults, low incidence in middle aged adults, and increasing incidence with age among older adults (Hjalgrim and Jarrett, 2020). Overall, incidence is approximately 25-45% higher in men compared to women (Cook *et al.*, 2009, Radkiewicz *et al.*, 2017). Age-standardized incidence rates per 100,000 people aged 35 years and above in the UK diagnosed between 2013 and 2017 were 4.2 for men and 2.4 for women (Table 2.1). Risk factors associated with HL include infections, particularly EBV and HIV, autoimmune

diseases, immune suppression, and family history (Hjalgrim and Jarrett, 2020). Environmental risk factors include exposure to UV light and tobacco (Hjalgrim and Jarrett, 2020).

## Non-Hodgkin Lymphoma [C81 & C85]

Non-Hodgkin lymphoma (NHL) comprises a diverse group of cancers affecting lymphocytes, and is the most common haematological malignancy, ranking as the sixth most diagnosed cancer in the UK (Armitage *et al.*, 2017, Smedby *et al.*, 2018). The incidence of NHL is higher in men, with a reported 40%-50% greater risk compared to women (Cook *et al.*, 2009, Miranda-Filho *et al.*, 2019, Radkiewicz *et al.*, 2017). In 2016, the age-standardised incidence rates per 100,000 individuals were 9.7 for women and 13.5 for men (Miranda-Filho *et al.*, 2019).

Risk factors are associated with the development of NHL include immune deficiencies and disorders, such as rheumatoid arthritis, coeliac disease, systemic lupus erythematosus, and Sjögren's syndrome (Boffetta, 2011, Smedby *et al.*, 2018). The use of medications, particularly those used for the treatment of autoimmune diseases, have also been implicated (Boffetta, 2011). Infections like HIV/AIDS, herpes virus, Epstein-Barr virus, and Helicobacter pylori are established contributors (Boffetta, 2011). Other factors include type II diabetes, family history, genetic conditions, and occupational or environmental exposure to carcinogens, including personal use of hair dyes (Boffetta, 2011).

## Multiple Myeloma [C90 & C88]

Multiple myeloma (hereafter referred to as myeloma) is a cancer of plasma cells within the bone marrow and represents the third most common haematological malignancy worldwide, following non-Hodgkin lymphoma and leukaemia (Huang *et al.*, 2022a). Myeloma exhibits a higher incidence in men compared to women, with a reported 50-60% greater incidence in men in countries with a high development index (Cook *et al.*, 2009, Huang *et al.*, 2022a).

Additionally, it is reported the incidence of myeloma has among adults 50 years of age and above, particularly men. Several risk factors associated with myeloma include overweight and obesity, genetic predispositions such as monoclonal gammopathy of undetermined significance, a family history of the disease, and autoimmune conditions like pernicious anaemia, and occupational factors (Huang *et al.*, 2022a). The cancer is also more common in Black populations and has been linked to occupational exposures to carcinogens, such as those encountered by firefighters, hairdressers, and individuals working with pesticides, herbicides, or other hazardous agents (Sergentanis *et al.*, 2015). The observed higher risk in men may be attributed to greater exposure to environmental and occupational risk factors (Huang *et al.*, 2022a).

## Leukaemia [C91-C95]

Leukaemia refers to a group of four main subtypes of blood cell cancers, all of which exhibit a higher incidence in men than in women among adults aged over 20 years (Daltveit *et al.*, 2025, Roman *et al.*, 2018). The two predominant subtypes in adults are acute myeloid leukaemia (AML) and chronic lymphocytic leukaemia (CLL). Globally, the incidence of leukaemia varies, with the highest rates observed in countries with a high Human Development Index, including the UK (Daltveit *et al.*, 2025). Between 2013 and 2017, ASIRs per 100,000 person-years were 22.9 for men and 13.0 for women (Table 2.2). In the UK in 2022, AML represented nearly 30% of leukaemia cases; ASIRs per 100,00 person-years were 3.1 for women and 4.2 for men (Daltveit *et al.*, 2025). CLL accounted for over 40% of all leukaemia cases; ASIRs per 100,000 person years 6.1 for men and 3.5 for women (Daltveit *et al.*, 2025). Notably, overall sex differences in leukaemia incidence among adults increase with age, becoming particularly pronounced after 65 years (Miranda-Filho *et al.*, 2018).

Several risk factors are associated with leukaemia, including smoking, exposure to carcinogenic agents, certain blood disorders, exposure to radiation and chemotherapy

(particularly for individuals treated for blood disorders), infections (e.g., human T-cell lymphotropic virus, herpes virus), specific genetic syndromes, and a family history of leukaemia (Roman *et al.*, 2018). Carcinogenic exposures encompass occupational hazards, such as work in the leather or printing industries, as well as exposure to benzene, Agent Orange, and radon (Roman *et al.*, 2018).

## 2.5 Summary

The aim of this thesis chapter was to review the sex differences in cancer incidence, highlighting the disparities observed across various cancer sites and across the extrinsic and intrinsic risk factors understood to contribute to them. Substantial epidemiological evidence indicates that, for most cancer sites common to both sexes, incidence rates are higher in men compared to women, with relative risks ranging from 1.5- to 3- fold. Cancers of the bladder, oesophagus, oral cavity, and liver have up to threefold higher risk in men compared to women. Cancers of the gallbladder, anus, and thyroid and lung adenocarcinoma conversely have higher risk in women compared to men.

40% of cancer incidence in the UK is attributable to known modifiable risk factors (Brown *et al.*, 2018). Temporal trends in cancer incidence differ by sex in ways that broadly mirror sex-specific patterns of risk factor exposure, and it is hypothesised that differential risk factor distributions contribute substantially to observed sex differences in cancer incidence. Health-related factors—smoking, overweight and obesity, and alcohol consumption—are important contributors for risk of many cancers and are hypothesized to contribute to the sex differences observed for many cancer sites given their predominantly higher prevalence in men compared to women. Many of the studies referred to in this literature review have described sex differences in cancer incidence deriving from national cancer registry data, which characterise population-level patterns but provide limited insight into the individual-levels factors that may contribute to them. A prospective cohort study evaluating sex differences in cancer incidence for 21 cancer sites assessed the adjusted effect of sex on cancer incidence with risk factors selected based on their univariate associations with cancer endpoints (Jackson *et al.*, 2022). They included adjustments for smoking, BMI, and alcohol for most cancer sites, and noted significant sex differences remained. However, the authors noted that their analyses could not exclude the influence of unmeasured or residual

confounding from these risk factors. Noting this, a consideration for this doctoral work has been to carefully evaluate contributions from these three risk factors in examining sex differences in cancer incidence, especially for cancer sites for which they are well-established risk factors.

Height, as an intrinsic risk factor and a proxy for increased cell number, is an important but not sole contributor to sex differences in cancer incidence for many cancer sites (Fu *et al.*, 2020b, Jackson *et al.*, 2022, Walter *et al.*, 2013).

As mentioned previously, the aim of this thesis is the assessment of whether and to what extent adjustment for risk factors contribute to observed sex differences in cancer incidence. Based on the literature review, extrinsic factors comprising sociodemographic characteristics (ethnicity and educational qualifications) and health-related behaviours (smoking and alcohol consumption), while intrinsic biological factors including body mass index (BMI) and height, were deemed important to consider when evaluating sex differences, reflecting their established relevance to cancer risk and sex differences in distributions. Educational attainment was considered as it contributes to the effects of health-related behaviours on cancer risk, particularly smoking, cancer screening behaviours, and, among women, reproductive factors. Ethnicity was considered as it may reflect underlying variations in distributions of environmental and health-related factors between groups. In addition, several cancer site-specific risk factors hypothesized to contribute to sex differences were highlighted in this literature review, which formed the second consideration for this doctoral work, to evaluate their contributions to sex differences in cancer incidence.

# Chapter 3 UK Biobank Cohort: Design, Variables, and Principal Statistical Approach

Here, details of the data and methods applied for the statistical analyses of chapters 4-6 have been given, including (1) the UK Biobank cohort design and baseline assessment; (2) analytic cohort characteristics, (3) participant exclusions; (4) exposure categorisation; (5) ascertainment of cancer incidence; (6) cancer endpoints selected for analyses, (7) statistical power calculations, (8) principal Cox regression modelling approach adopted throughout analyses, and (9) the selected risk factor variables derivation and categorisation. Chapter specific statistical models and sensitivity analyses have been subsequently detailed in each chapter respectively.

## 3.1 UK Biobank Cohort Design and Baseline Assessment

The UK Biobank is an ongoing cohort study of half a million participants (aged 37-73 years at recruitment) from the United Kingdom, registered with the UK National Health Service (NHS, hereafter) (Collins, 2012, Fry *et al.*, 2017). Between 2006-2010, approximately 9.2 million NHS-registered adults aged between 37-73 residing within ~25 miles of one the 22 assessment centres across England, Wales and Scotland were invited to volunteer (Collins, 2012, Fry *et al.*, 2017). Around 502,000 participants volunteered, yielding a 5.5% response rate (of which, 6.4% were women and 5.1% were men) (Fry *et al.*, 2017, UK Biobank Coordinating Centre, 2006). The study was conducted in accordance with the Helsinki Declaration of 1975 as revised in 1983, and was approved by the North West Multi-Centre

Research Ethics Committee (UK) (06/MRE08/65), the Patient Information Advisory Group (England and Wales), the National Information Governance Board for Health and Social Care (England and Wales), and the Community Health Index Advisory Group (Scotland) (UK Biobank, 2007).

Participants joined the study by completing baseline touchscreen questionnaires, providing information on socio-demographics, health-related behaviours (including diet, physical activity, smoking, and alcohol use), family and personal medical history, among several other health-related factors (Fry *et al.*, 2017). They volunteered for measurement of anthropometric and biological factors—such as height, weight, bioimpedance derived anthropometric measurement, blood pressure, lung function tests—and provided blood, urine, and saliva samples for long-term storage. They additionally consented for their health to be followed up through linkage to their electronic medical records. Baseline assessment visits lasted from 90-180 minutes and were conducted by trained staff using the cohort's published standard operating procedures and manuals, ensuring consistency across recruitment centres (Fry *et al.*, 2017) .

All UK Biobank variables referenced in this thesis are identified by their corresponding data-field IDs (UK Biobank, 2017). To allow ease of cross-reference, these data-fields are publicly documented and may be accessed via the UKB Showcase (UK Biobank, 2017).

## 3.2 Analytic Cohort Characteristics

The UK Biobank study cohort included 470,771 participants aged 37 to 73 years at recruitment, 253,344 (53.8%) women and 217,427 (46.2%) men (Table 3.1). On average, women were aged 56 (SD 8) and men were aged 56 (SD 8) at recruitment. Mean follow-up was an average of 10.5 (SD 2.2) years (10.6 (SD 2.1) years for women and 10.4 (SD 2.4) years for men) and participants contributed 49.4 (100,000s) person-years of observation, including 26.9 among women and 22.4 among men. 32,315 incident cancers across the

shared anatomical cancer sites included in the present analyses were identified: 18,774 cancers (58.1%) in women and 13,541 (41.9%) in men (further detailed in Section 3.6 and Table 3.2).

The largest proportion of participants joining the study was from North-West England (15.6%), Yorkshire and the Humber (14.8%) and London (13.6%). The regional distribution of participants was broadly similar between women and men, with only small differences observed across recruitment regions. The largest absolute sex differences were observed for London, where women comprised a slightly greater proportion of participants than men (13.3% men and 14.0% women), and the West Midlands, where men comprised a slightly greater proportion than women (9.5% men and 8.4% women). The distribution of participants across deprivation quintiles was similar between women and men.

Mean BMI was 27.1 (SD 5.2) $kg/m^2$ in women and 27.9 (SD 4.3) $kg/m^2$ in men. More women had BMI <25 $kg/m^2$ compared with men (24.6% men and 39.2% women). Men were more commonly in the overweight (25-29.9 $kg/m^2$; 49.5% men and 36.9% women) and BMI 30-34.9 $kg/m^2$ (20.0% men and 15.8% women) categories. The proportion of participants with BMI ≥35 $kg/m^2$ was higher for women compared with men (5.9% men and 8.0% women). Mean height was 175.6 (SD 6.8) cm in men and 162.5 (SD 6.3) cm in women.

Qualification distributions were broadly similar between sexes, although men more commonly reported degree-level or professional qualifications than women (62.6% men and 56.6% women). Women more frequently reported O level/GCSE/CSE qualifications or equivalent (13.5% men and 19.2% women), whereas the proportion reporting no listed qualifications was similar between sexes (16.9% men and 16.5% women). Missing qualification data were low overall and comparable between sexes (1.9% men and 1.8% women).

The cohort was predominantly White, with 94.1% of both women and men identifying as White. Smaller proportions of participants identified as Asian or Asian British (2.6% men and 2.1% women), Black or Black British (1.5% men and 1.7% women), mixed race (0.5% men and 0.7% women), and other ethnic groups (1.0% men and 0.9% women). Missing ethnicity data were low overall and slightly higher among men compared with women (0.6% men and 0.4% women).

Smoking status and smoking intensity distributions differed between sexes. A higher proportion of women reported never smoking compared with men (50.0% men and 59.6% women). Men reported heavier smoking histories, including former smoking of ≥15 cigarettes per day (19.6% men and 12.4% women) and current smoking of ≥15 cigarettes per day (5.0% men and 3.5% women). Women reported slightly higher former smoking history of <15 cigarettes per day (5.3% men and 6.6% women). The proportion of current smokers reporting <15 cigarettes per day was low for both sexes (2.8% men and 3.4% women). Missing smoking data were also similar between men and women (0.6% men and 0.5% women).

Alcohol consumption distributions also differed between sexes. Women more commonly reported lower alcohol intake, including consumption of <1 g/day (6.6% men and 15.1% women) and ≥1-<10 g/day (22.0% men and 39.2% women). Men more frequently reported higher alcohol intake categories, particularly consumption of ≥20-<30 g/day (16.9% men and 8.6% women) and ≥30 g/day (26.4% men and 5.0% women). The proportion of non-drinkers was also higher among women than men (6.2% men and 9.4% women). Missing alcohol consumption data were low overall and slightly higher among women (0.5% men and 0.8% women).

Physical activity distributions were broadly similar between sexes. Moderate physical activity levels (≥10-<50 METS per week) were the most commonly reported category in both

men and women (41.5% and 39.9%, respectively). Men were slightly more likely to report high physical activity levels (≥50 METS per week; 19.8% men and 15.7% women) and low physical activity levels (0-<10 METS per week; 19.9% men and 18.2% women). Women had a greater proportion of missing physical activity data compared with men (18.7% men and 26.1% women).

Details of variable coding for each of these categories are further described in section 3.9.

**Table 3.1 Baseline distribution of covariates and details of follow-up in the UK Biobank.**

| | Women | Men | Total |
|---|---|---|---|
| **Cohort size** | 253,344 | 217,427 | 470,771 |
| **Follow-up for cancer incidence** | | | |
| Number of incident cancers | 18,774 | 13,541 | 32,315 |
| Person-years of follow up, (100,000s) | 26.9 | 22.4 | 49.4 |
| Mean follow-up time [SD] | 10.6 [2.1] | 10.4 [SD 2.4] | 10.5 [2.2] |
| Median follow-up time [IQR] | 10.9 [10.1-11.7] | 10.9 [9.9-11.6] | 10.9 [10.1-11.7] |
| **Age, years** | | | |
| Mean age at recruitment [SD] | 56.1 [8.0] | 55.6 [8.2] | 56.3 [8.1] |
| <45 | 26,967 (10.6) | 23,412 (10.8) | 50,379 (10.7) |
| 45-49 | 35,285 (13.9) | 28,497 (13.1) | 63,782 (13.5) |
| 50-54 | 40,712 (16.1) | 31,991 (14.7) | 72,703 (15.5) |
| 55-59 | 47,049 (18.6) | 38,334 (17.6) | 85,383 (18.1) |
| 60-64 | 60,034 (23.7) | 52,002 (23.9) | 112,036 (23.8) |
| >65 | 43,297 (17.1) | 43,191 (19.9) | 86,488 (18.4) |
| **Region** | | | |
| London | 35,512 (14.0) | 28,905 (13.3) | 64,417 (13.6) |
| North-West England | 38,656 (15.3) | 34,894 (16.1) | 73,550 (15.6) |
| North-Eastern England | 29,488 (11.6) | 25,358 (11.7) | 54,846 (11.7) |
| Yorkshire and the Humber | 37,506 (14.8) | 32,334 (14.9) | 69,840 (14.8) |
| West Midlands | 21,284 (8.4) | 20,695 (9.5) | 41,979 (8.9) |
| East Midlands | 17,194 (6.8) | 14,690 (6.8) | 31,884 (6.7) |
| South-East England | 22,419 (8.9) | 18,336 (8.4) | 40,755 (8.6) |
| South-West England | 22,314 (8.8) | 18,111 (8.3) | 40,425 (8.5) |
| Wales | 10,431 (4.1) | 8,996 (4.1) | 19,427 (4.1) |
| Scotland | 18,540 (7.3) | 15,108 (7.0) | 33,648 (7.1) |
| **Deprivation Index, in quintiles** | | | |
| Most affluent | 50,544 (19.9) | 43,884 (20.2) | 94,428 (20.0) |
| 2$^{nd}$ | 50,755 (20.0) | 43,411 (19.9) | 94,166 (20.0) |
| 3$^{rd}$ | 51,320 (20.4) | 42,785 (19.7) | 94,105 (20.0) |
| 4$^{th}$ | 51,234 (20.2) | 42,724 (19.7) | 93,958 (20.0) |
| Most deprived | 49,197 (19.4) | 44,340 (20.4) | 93,537 (19.8) |
| Unknown* | 294 (0.1) | 283 (0.1) | 577 (0.1) |
| **BMI, kg/m$^2$** | | | |
| Mean BMI [SD] | 27.1 [5.2] | 27.9 [4.3] | 27.4 [4.8] |
| <25 | 99,284 (39.2) | 53,579 (24.6) | 152,863 (32.5) |
| 25-29.9 | 93,652 (36.9) | 107,730 (49.5) | 201,382 (42.8) |
| 30-34.9 | 40,050 (15.8) | 43,377 (20.0) | 83,427 (17.7) |
| >=35 | 20,358 (8.0) | 12,741 (5.9) | 33,099 (7.0) |
| **Height, cm** | | | |
| Mean height [SD] | 162.5 [6.3] | 175.6 [6.8] | 168.5 [9.3] |
| **Qualifications** | | | |

| | | | |
|---|---|---|---|
| Prof Q/NVQ/HND/HNC/Degree or other professional qualification | 143,289 (56.6) | 136,211 (62.6) | 279,500 (59.4) |
| A Levels/or equivalent | 14,694 (5.8) | 10,922 (5.0) | 25,616 (5.4) |
| O level/GCSE/CSE/or equivalent | 48,762 (19.2) | 29,330 (13.5) | 78,092 (16.6) |
| None of the above | 41,864 (16.5) | 36,663 (16.9) | 78,527 (16.7) |
| Unknown* | 4,735 (1.8) | 4,301 (1.9) | 9,036 (1.9) |
| **Ethnicity** | | | |
| White | 238,441 (94.1) | 204,496 (94.1) | 442,937 (94.1) |
| Mixed race | 1,738 (0.7) | 1,066 (0.5) | 2,804 (0.6) |
| Asian or Asian British | 5,303 (2.1) | 5,551 (2.6) | 10,854 (2.2) |
| Black or Black British | 4,401 (1.7) | 3,213 (1.5) | 7,614 (1.6) |
| Other | 2,432 (1.0) | 1,876 (0.9) | 4,308 (0.9) |
| Unknown* | 1,029 (0.4) | 1,225 (0.6) | 2,254 (0.5) |
| **Smoking status and intensity, cigarettes per day** | | | |
| Never | 151,008 (59.6) | 106,404 (50.0) | 257,412 (54.7) |
| Former, <15 | 16,717 (6.6) | 11,565 (5.3) | 28,282 (6.0) |
| Former, >=15 | 31,333 (12.4) | 42,615 (19.6) | 73,948 (15.7) |
| Former, unknown | 30,434 (12.0) | 28,416 (13.1) | 58,850 (12.5) |
| Current, <15 | 8,270 (3.4) | 6,058 (2.8) | 14,328 (3.0) |
| Current, >=15 | 8,850 (3.5) | 10,854 (5.0) | 19,704 (4.1) |
| Current, unknown | 5,412 (2.1) | 10,183 (4.7) | 15,595 (3.3) |
| Unknown* | 1,320 (0.5) | 1,332 (0.6) | 2,652 (0.5) |
| **Alcohol consumption, grams per day** | | | |
| <1 | 38,298 (15.1) | 14,403 (6.6) | 52,701 (11.1) |
| >=1-<10 | 99,386 (39.2) | 47,807 (22) | 147,193 (31.2) |
| >=10-<20 | 55,341 (21.8) | 46,306 (21.3) | 101,647 (21.6) |
| >=20-<30 | 21,887 (8.6) | 36,715 (16.9) | 58,602 (12.4) |
| >=30+ | 12,637 (5.0) | 57,476 (26.4) | 70,113 (14.9) |
| Non-drinkers | 23,898 (9.4) | 13,573 (6.2) | 37,471 (7.9) |
| Unknown* | 1,904 (0.8) | 1,151 (0.5) | 3,055 (0.6) |
| **Physical activity, metabolic equivalents (METS) per week** | | | |
| Low (0-<10) | 46,164 (18.2) | 43,264 (19.9) | 89,428 (18.9) |
| Moderate (>=10-<50) | 101,089 (39.9) | 90,236 (41.5) | 191,325 (40.6) |
| High (>=50) | 39,741 (15.7) | 43,090 (19.8) | 82,831 (17.6) |
| Unknown* | 66,320 (26.1) | 40,816 (18.7) | 10,7136 (22.7) |

Results shown are N (%) unless otherwise specified
*This category included participants who responded either 'Do not know' or 'prefer not to say'.

## 3.3 Exclusion Criteria

The UK Biobank cohort dataset initially comprised 502,386 participants (Appendix Table 3A). Participants were excluded from this cohort if they had missing baseline information on key demographic or anthropometric variables required for pan-cancer analyses, including sex [data-field 31], region [data-field 54], age at recruitment [data-field 21022], standing height [data-field 50], or body mass index (BMI) [data-field 21001]. A total of 3,105 participants were excluded. In addition, participants with a prevalent cancer diagnosis at baseline were excluded (n=28,143), with the exception of non-melanoma skin cancer (ICD-10 C44). Non-melanoma skin cancer was retained due to its high incidence and incomplete ascertainment in cancer registries. Furthermore, 367 participants were excluded due to discordance between self-reported sex and genetically inferred sex (also refer below to section 3.4).

The primary analytical cohort comprised 470,771 participants and formed the basis for the principal analyses presented in Chapter 4, including both minimally-adjusted and multivariable-adjusted Cox proportional hazards models. For multivariable-adjusted models with cancer specific covariates, 51 participants with a history of partial or complete gastrectomy were additionally excluded (final analytical cohort n=470,720). For Chapter 5, which examined associations of body composition measures with cancer risk, an additional 7,955 participants with missing data at baseline on waist circumference, hip circumference, body fat and trunk fat measures were excluded (final analytical cohort n=462,675 participants). For Chapter 6, which examined lung adenocarcinoma in never smokers, participants who reported being current or former smokers at baseline were excluded (final analytical cohort n=257,390). The exclusion criteria are detailed in Appendix Table 3A.

## 3.4 Exposure: Sex

Participants' sex was acquired from their National Health Service (NHS) records at recruitment, but in some cases, this may have been updated based on self-reports (UKB data field 31). It was further checked whether participants' self-reported sex matched their genotyped sex (data field 22001 with two categories, female and male; available for 98.98% participants). 367 participants were excluded if they had differences between their sex as recorded on the NHS record or self-reported at recruitment and their sex as determined by UK Biobank genotyping analyses.

## 3.5 Ascertainment of Cancer Incidence

Cancer diagnoses were ascertained through linkage to national cancer and death registries, and the outcomes were defined as first incident cancer diagnosis or cancer first recorded in death certificates. All outcomes were defined according to the World Health Organization's International Statistical Classification of Diseases (ICD-10) (World Health Organization, 2019). Cancer data and death data in England and Wales are provided by NHS England, and in Scotland by the NHS Central Register and the National Records of Scotland (UK Biobank, 2023a, UK Biobank, 2023b). The person-years of follow-up were calculated from baseline assessment at recruitment until the first registration of cancer, date of death, loss, or end of follow up, whichever came first. The cancer data censoring dates were 31 December 2020 for England, 31 December 2016 for Wales, and 30 November 2021 for Scotland (UK Biobank, 2023a, UK Biobank, 2023b).

## 3.6 Cancer Endpoints and Case Counts

28 cancer endpoints (15 main sites and 13 subtypes) were considered for investigation of sex differences with at least 100 cancer cases for each (cases totalling for men and women), comprising the *pan-cancer* scope of prospective analyses undertaken in this

thesis (Table 3.2). These cancer endpoints and their respective ICD-10 codes were: cancers of the lip, oral cavity, and pharynx [C00-C14] (oral cavity, hereafter), oesophagus [C15] with oesophageal adenocarcinoma and oesophageal squamous cell carcinoma, stomach [C16] with gastric cardia [C16.0] and gastric non-cardia [C16.1-C16.6], colorectum [C18-C20] with colon [C18] and rectum [C19-C20], anus [C21], liver [C22], gallbladder [C23], pancreas [C25], lung [C34] with lung adenocarcinoma, lung small cell carcinoma and lung squamous cell carcinoma, malignant melanoma [C43] (melanoma, hereafter), breast [C50], kidney [C64-C65], bladder [C67], thyroid [C73], and lymphatic and hematopoietic sites [C81-C96], with Hodgkin lymphoma [C81], non-Hodgkin lymphoma (NHL) [C82-C85], multiple myeloma [C88 & C90], and leukaemia [C91-C95].

32,315 incident cancers across the shared anatomical sites included in the present analysis were identified: 18,774 cancers (58.1%) in women and 13,541 (41.9%) in men (Table 3.2). Of the cancer cases identified, the most frequently diagnosed cancers were breast (8,792 cases), colorectum (5,388 cases), lymphatic and haematopoietic cancers (4,317 cases), lung (3,664 cases), and melanoma (2,706 cases). The five most common site-specific cancers in men were colon, lung, melanoma, rectum, and non-Hodgkin lymphoma. In women, they were breast, lung, colon, melanoma, and non-Hodgkin lymphoma. Additionally, several cancers had sex differences in their case distribution. Men had a higher frequency of oesophageal adenocarcinoma (83.8% of all cases), bladder cancer (77.8%), gastric cardia (78.2%), and liver cancer (66.6%). Comparatively, women had a higher frequency of cancers of the breast (99.4% of all cases), thyroid (73.7%), anus (65.8%), and gallbladder (63.8%).

Breast cancer was included because, although this cancer is rare in men, there are some aetiological risk factors shared between men and women (Anderson *et al.*, 2010, Brinton *et al.*, 2015, Brinton *et al.*, 2008). Inclusion of central nervous system tumours was

considered but decided against, as they comprise a heterogenous group with distinct aetiology, risk factors and diagnostic characteristics (Louis *et al.*, 2021, Näslund *et al.*, 2023, Salari *et al.*, 2023). Oesophageal and lung histological subtypes were identified in reference to ICD-O-3 codes (Pirie *et al.*, 2016, Ruhl *et al.*, 2022, Watts *et al.*, 2021, Wong *et al.*, 2020, World Health Organization, 2013). Additionally, appendix tables list the ICD-10 codes for cancer sites examined for pan-cancer analyses (Chapter 4) (Appendix Table 3B), ICD-10 codes comprising cancers of the oral cavity (Appendix Table 3C), and the histology codes for oesophageal and lung cancer subtypes (Appendix Table 3D).

**Table 3.2 Cancer case counts for the 28 selected endpoints in the UK Biobank analytic cohort of 470,771 participants.**

| | Women (%) | Men (%) | Total cases |
|---|---|---|---|
| **Oral cavity [C00-C14]** | 324 (33.8) | 634 (66.2) | 958 |
| **Oesophagus [C15]** | 249 (26.0) | 707 (74.0) | 956 |
| **Adenocarcinoma** | 110 (16.2) | 571 (83.8) | 681 |
| **Squamous cell carcinoma** | 128 (52.5) | 116 (47.5) | 244 |
| **Stomach [C16]** | 203 (31.3) | 446 (68.7) | 649 |
| **Gastric cardia [C16.0]** | 62 (21.8) | 223 (78.2) | 285 |
| **Gastric non-cardia [C16.1-C16.6]** | 77 (39.9) | 116 (60.1) | 193 |
| **Colorectum [C18-C20]** | 2316 (43.0) | 3072 (57.0) | 5388 |
| **Colon [C18]** | 1667 (46.6) | 1909 (53.4) | 3576 |
| **Rectum [C19-C20]** | 663 (36.0) | 1178 (64.0) | 1841 |
| **Anus [C21]** | 127 (65.8) | 66 (34.2) | 193 |
| **Liver [C22]** | 194 (33.4) | 387 (66.6) | 581 |
| **Gallbladder [C23]** | 74 (63.8) | 42 (36.2) | 116 |
| **Pancreas [C25]** | 531 (45.9) | 627 (54.1) | 1158 |
| **Lung [C34]** | 1779 (48.6) | 1885 (51.4) | 3664 |
| **Adenocarcinoma** | 825 (54.2) | 697 (45.8) | 1522 |
| **Squamous cell carcinoma** | 293 (37.4) | 490 (62.6) | 783 |
| **Small cell carcinoma** | 156 (45.0) | 191 (55.0) | 347 |
| **Melanoma [C43]** | 1335 (49.3) | 1371 (50.7) | 2706 |
| **Breast [C50]** | 8737 (99.4) | 55 (0.6) | 8792 |
| **Kidney [C64-C65]** | 492 (35.7) | 885 (64.3) | 1377 |

| | | | |
|---|---|---|---|
| **Bladder [C67]** | 237 (22.2) | 830 (77.8) | 1067 |
| **Thyroid [C73]** | 292 (73.7) | 104 (26.3) | 396 |
| **Lymphatic and haematopoietic cancers [C81-C96]** | 1886 (43.7) | 2431 (56.3) | 4317 |
| **Hodgkin lymphoma [C81]** | 63 (47.7) | 69 (52.3) | 132 |
| **Non-Hodgkin lymphoma [C82-C85]** | 931 (46.4) | 1074 (53.6) | 2005 |
| **Multiple myeloma [C90 & C88]** | 383 (43.1) | 505 (56.9) | 888 |
| **Leukaemia [C91-C95]** | 500 (39.5) | 766 (60.5) | 1266 |

## 3.7 Statistical Power Calculations

Statistical power was assessed for each cancer endpoint using Schoenfeld's event-based approximation for Cox proportional hazards models (Hsieh and Lavori, 2000, Schoenfeld, 1983) (Table 3.3). For each endpoint, the number of incident cancer cases was used as the number of events. Calculations were done assuming a binary exposure of sex, the observed cohort sex distribution of 217,427 men (46.2%) and 253,344 women (53.8%), two-sided $\alpha$=0.05 and 80% power. For each cancer endpoint, the minimum detectable hazard ratio was estimated, and values greater than 1 indicated the magnitude of association detectable with 80% power. The corresponding detectable inverse association was given by the reciprocal of this value.

Statistical power varied substantially across cancer endpoints according to the number of incident cancer cases observed, with more common cancers having providing power to detect smaller sex differences in cancer risk than rarer cancer endpoints. Breast and colorectal cancer had sufficient power to detect modest sex differences in cancer risk (HR <1.10), whereas rarer cancers, including gallbladder cancer and Hodgkin lymphoma, required larger effect sizes (HR >1.60) to achieve equivalent statistical power. Consequently, estimates for less common cancers were interpreted with greater caution because of wider confidence intervals and reduced statistical precision.

**Table 3.3 Statistical power calculations for the 28 selected cancer endpoints in the UK Biobank analytic cohort.**

| | Total cases | Minimum detectable HR at 80% power |
|---|---|---|
| **Oral cavity [C00–C14]** | 958 | 1.20 |
| **Oesophagus [C15]** | 956 | 1.20 |
| **Adenocarcinoma** | 681 | 1.24 |
| **Squamous cell carcinoma** | 244 | 1.43 |
| **Stomach [C16]** | 649 | 1.25 |
| **Gastric cardia [C16.0]** | 285 | 1.39 |
| **Gastric non-cardia [C16.1–C16.6]** | 193 | 1.50 |
| **Colorectum [C18–C20]** | 5,388 | 1.08 |
| **Colon [C18]** | 3,576 | 1.10 |
| **Rectum [C19–C20]** | 1,841 | 1.14 |
| **Anus [C21]** | 193 | 1.50 |
| **Liver [C22]** | 581 | 1.26 |
| **Gallbladder [C23]** | 116 | 1.68 |
| **Pancreas [C25]** | 1,158 | 1.18 |
| **Lung [C34]** | 3,664 | 1.10 |
| **Adenocarcinoma** | 1,522 | 1.15 |
| **Squamous cell carcinoma** | 783 | 1.22 |
| **Small cell carcinoma** | 347 | 1.35 |
| **Melanoma [C43]** | 2,706 | 1.11 |
| **Breast [C50]** | 8,792 | 1.06 |
| **Kidney [C64–C65]** | 1,377 | 1.16 |
| **Bladder [C67]** | 1,067 | 1.19 |
| **Thyroid [C73]** | 396 | 1.33 |
| **Lymphatic and haematopoietic cancers [C81–C96]** | 4,317 | 1.09 |
| **Hodgkin lymphoma [C81]** | 132 | 1.63 |
| **Non-Hodgkin lymphoma [C82–C85]** | 2,005 | 1.13 |
| **Multiple myeloma [C90 & C88]** | 888 | 1.21 |
| **Leukaemia [C91–C95]** | 1,266 | 1.17 |

Minimum detectable hazard ratios (HRs) were calculated using Schoenfeld's approximation for Cox proportional hazards models, assuming two-sided $\alpha = 0.05$, 80% power, and the observed cohort sex distribution. For inverse associations, the corresponding detectable HR is the reciprocal of the value shown.

## 3.8 Principal Cox Regression Modelling Approach

Cox proportional hazards (PH) regression models with attained age as the underlying timescale were used to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) for the risk of cancer in men compared to women, treating each cancer as a different endpoint.

Using age as the underlying timescale variable defines risk sets among individuals of the same chronological age and allows the baseline hazard to represent the age-specific incidence function without imposing parametric assumptions on the effect of age (Griffin *et al.*, 2012, Peto *et al.*, 1976, Thiébaut and Bénichou, 2004) . Thus, it is ensured that comparisons of exposure effects are made at comparable ages and residual confounding or bias that can arise from the effect of age is avoided (Griffin *et al.*, 2012, Peto *et al.*, 1976, Thiébaut and Bénichou, 2004).

First, a minimally-adjusted Cox model was used with stratification by age group at recruitment (< 45, 45 to <50, 50 to <55, 55 to <60, 60 to <65, and ≥65 years), Townsend index of deprivation (quintiles; a score that incorporates census area data for employment, car ownership, home ownership, and household overcrowding) (Townsend *et al.*, 1988), and region (ten regions: London, North-West, North-East, Yorkshire and Humber, West Midlands, East Midlands, South-East, South-West, Wales, and Scotland).

Second, a baseline multivariable-adjusted Cox model was used, which additionally adjusted for qualification status (college or university degree/vocational qualification, A levels or equivalent (national examination at 17-18 years), O levels or equivalent (national examination at 16 years), and other/unknown qualifications), ethnicity (Asian or Asian British, Black or Black British, Mixed, Other, and White), height (continuous, measured in cm), body mass index (BMI; calculated as weight in kilograms divided by height in metres squared, <25, 25 to <30, 30 to <35, and ≥35 $kg/m^2$), smoking (categorised by a composite

variable that combined smoking intensity and status; never, former (<15 and ≥15 cigarettes/day), and current (<15 and ≥15 cigarettes/day)), and alcohol intake (categorised based on intake in grams per day; never drinkers, <1, 1 to <10, 10 to <20, 20 to <30 and 30+ grams/day). Missing categories were generated for the following variables: qualifications (for which, 0.9% participants had missing data), ethnicity (0.5% missing), smoking (0.5% missing), and alcohol (0.6% missing). Categorical adjustment variables were used throughout to minimise any undue influence of extreme values, except for height for which a continuous variable was used because of the substantial non-overlap between women and men for this variable and because it was measured under standardised conditions and therefore there should be no erroneous extreme values. Details of variable coding have been given in the next section (Section 3.9).

The pan-cancer analyses additionally adjusted for cancer-site specific risk factors, which were added to the multivariable-adjusted Cox models (described in the methods section of Chapter 4). Analyses for Chapters 5 and 6 also applied the baseline (described here) and cancer-specific multivariable-adjusted Cox modelling approaches (Chapter 4), along with chapter specific methods fitting to the aim of the analyses.

In Chapters 4 and 5, changes in sex-specific cancer associations across sequentially adjusted models were assessed on the natural logarithm of the hazard ratio (HR; ln(HR)) scale. The percentage change in the log hazard ratio (%Δln(HR)) was used to quantify the extent to which adjustment for covariates changed the estimated differences by sex and was calculated as $[\ln(\mathrm{HR}_{\mathrm{unadjusted}}) - \ln(\mathrm{HR}_{\mathrm{adjusted}})]/\ln(\mathrm{HR}_{\mathrm{unadjusted}}) \times 100]$. Positive values indicated attenuation of the association towards the null, whereas negative values indicated strengthening of the association after adjustment. The approach was used descriptively to assess the impact of covariate adjustment and was interpreted alongside the corresponding HRs (95% confidence intervals), and case counts. In Chapter 4,

%Δln(HR) was used to compare minimally adjusted and fully adjusted models across the 28 cancer endpoints included in the primary analyses. In Chapter 5, %Δln(HR) was additionally used to assess attenuation of sex differences in oesophageal adenocarcinoma and kidney cancer after sequential adjustment for individual adiposity traits relative to a reference model without adiposity adjustment.

For Chapter 5 analyses, Z-scores were additionally examined descriptively to compare the standardized statistical evidence for sex-cancer associations across sequentially adjusted models. They are calculated as a ratio of the ln(HR) to its standard error [Z= $\beta/SE(\beta)$]. Smaller absolute Z-scores reflected weaker statistical evidence against the null hypothesis. The Z-scores were not interpreted as measures of effect size and were considered alongside HRs (95% confidence intervals), and the %Δ ln(HR).

All analyses were performed using STATA (version 18 and 19) and results were plotted using R and RStudio (Posit Team, 2025, StataCorp, 2025).

# 3.9 Variable Coding for Baseline Multivariable-adjusted Analyses

## 3.9.1 Age at Recruitment

Age at recruitment (data field 21022) is a derived variable based on a participant's date of birth and date of attending an initial assessment at the recruitment centres. Age was categorized into 6 groups: < 45, 45 to <50, 50 to <55, 55 to <60, 60 to <65, and ≥65 years.

## 3.9.2 Region

Region (data field 54) refers to the respective baseline assessment centres, which were based in Barts, Hounslow, Croydon, Swansea, Wrexham, Cardiff, Stockport, Manchester, Liverpool, Bury, Newcastle upon Tyne, Middlesbrough, Leeds, Sheffield, Stoke, Birmingham, Nottingham, Oxford, Reading, Bristol, Glasgow, and Edinburgh. They were

grouped into 10 regions: London (Barts, Hounslow, and Croydon), Wales (Wrexham, Swansea, and Cardiff), North-West (Stockport, Manchester, Liverpool, and Bury), North-East (Newcastle and Middlesbrough), Yorkshire and Humber (Leeds and Sheffield), West Midlands (Stoke and Birmingham), East Midlands (Nottingham), South-East (Oxford and Reading), South-West (Bristol), and Scotland (Glasgow and Edinburgh).

### 3.9.3 Townsend Deprivation Index

Material deprivation is measured by the Townsend Deprivation Index (TDI); a score that incorporates census area data for employment, car ownership, home ownership, and household overcrowding (Townsend *et al.*, 1988). TDI (data field 22189) was derived prior to recruitment using preceding national census output area, and participants were assigned a score corresponding to the output area in which their postcode was located. For the analysis, TDI was categorised using quintiles, and participants without TDI information were categorized as 'unknown' (0.1%).

### 3.9.4 Ethnicity

Ethnicity (data field 21000) was assessed using one question on ethnicity 'What is your ethnic group?' with options for 'White', 'Mixed', 'Asian or Asian British', 'Black or Black British', 'Chinese', 'Other ethnic group', 'Do not know' or 'Prefer not to answer' and four supporting questions within each group. 'White' (data field 1667) was defined as either answering white, British, Irish or any other white background; 'Asian or Asian British' was defined as Asian or Asian British, Chinese, Indian, Pakistani, Bangladeshi, or any other Asian background; 'Black or Black British' (data field 3690) was defined as Black or Black British, Caribbean, African or any other Black background; 'Mixed Race or other' (data field 3819) was defined as reporting either any other ethnic groups or mixed ethnicity; participants with missing data, reporting they did not know or preferred not to answer, were grouped as 'unknown' (0.6%).

### 3.9.5 Qualifications

Highest qualifications (data field 6139) were assessed within the touchscreen questionnaire 'Which of the following qualifications do you have?', with the options 'College or university degree', 'A levels/AS levels or equivalent', 'O levels/GCSEs or equivalent', 'CSEs or equivalent', 'NVQ or HND or HNC or equivalent', 'Other professional qualifications e.g: nursing, teaching', 'None of the above' and 'Prefer not to answer'. Qualification was then categorized into three groups: the highest qualifications included 'College or university degree', 'NVQ or HND or HNC or equivalent', 'Other professional qualifications e.g.: nursing, teaching' and are referred to as college or university degree/vocational qualification; the intermediate qualifications included 'A levels/AS levels or equivalent' and are referred to as national examination at age 17-18; and the lowest level of qualifications included 'O levels/GCSEs or equivalent', 'CSEs or equivalent' and are referred to as national examination at age 16. Participants who had answered 'None of the above' or 'prefer not to answer' or who had not answered the question were categorized as 'unknown/other' (18.6%).

### 3.9.6 Height

Standing height (data field 50) was measured at baseline assessment with a Seca 240 cm height measure using a standardized protocol. Where height measurements were missing participants were categorized into an 'unknown' group (0.5%).

### 3.9.7 Body mass index (BMI)

Body mass index (data field 23104) was calculated as weight in kilograms divided by height in meters squared. Participant weight was measured at baseline using a Tanita BC418MA body composition analyser as part of Bioimpedance measurement. For participants that did not participate in a Bioimpedance analysis, body weight was based on measurement

using a standard scale, calculated using weight in kilograms (data field 21002) divided by height in metres squared (data field 50). Where height and/or weight measurements were missing, participants were categorized into an 'unknown group' and dropped from the analytical cohort (0.6%). Details of how the UK Biobank bioimpedance analyses were carried out have been described in the UK Biobank resource 1421 (UK Biobank, 2011).

### 3.9.8 Smoking

Smoking status (data field 20116) was derived from two touchscreen questionnaires at baseline, 'Do you smoke tobacco now?' (data field 1239) with four options, 'Yes, on most or all days', 'Only occasionally', 'No' and 'Prefer not to answer', and 'In the past, how often have you smoked tobacco?' (data field 1249) with the options 'Smoked on most or all days', 'Smoked occasionally', 'Just tried once or twice', 'I have never smoked', and 'Prefer not to answer'. Number of cigarettes per day was obtained from the question 'About how many cigarettes do you smoke on average each day' (data field 3456), and participants could either give a number or answer, 'less than one', 'do not know' and prefer not to answer'. A composite variable combining smoking status and smoking intensity was created, categorising it into 7 groups of 'never smokers', 'former smokers <15 cigarettes/day', 'former smokers >=15 cigarettes/day', 'current smokers <15 cigarettes/day', 'current smokers >=15 cigarettes/day', 'current smokers, amount unknown', and an 'unknown' smoking status (0.6%).

For lung cancers, involuntary exposure to smoking and air pollution was also adjusted for in analyses restricted to never-smokers. Exposure to second hand smoke was measured using two variables (data field 1269 and 1279), and only in participants who self-reported as never smokers (n=257,412). They were asked 'At your home, how many hours per week are you exposed to other people's tobacco smoke? (participants could select from 0 to 168 hours, do not know, or prefer not to answer), and 'Outside of your home, how many hours

per week are you exposed to other people's tobacco smoke? (With options ranging from 0-168 hours, do not know, and prefer not to answer). 18,970 participants who identified as never smokers reported they were exposed to at least 1 or more hours per week of tobacco smoke at home, with 61.1% women in this category. 77,847 participants who identified as never smokers reported they were exposed to at least 1 or more hours per week of tobacco smoke outside home, with 46.1% women in this category. The multivariable-adjusted model for lung cancers adjusted for exposure to tobacco smoke at home and outside home in a continuous format. Residential air pollution was adjusted for using particulate matter air pollution (PM2.5) levels (data field 24006), estimated Air pollution estimates for the year 2010 were modelled for each address using a Land Use Regression (LUR) model developed as part of the European Study of Cohorts for Air Pollution Effects (ESCAPE) (Beelen *et al.*, 2014). Adjustment for these variables in lung cancer models has been discussed and described in chapters 4 and 6.

### 3.9.9 Alcohol

Alcohol drinking status (data field 20117) was derived at recruitment and participants were categorised as current drinkers, previous drinkers, never drinkers, and those who preferred not to answer. One question on alcohol intake frequency (data field 1558) was 'about how often do you drink alcohol?' categorised into 'daily or almost daily', 'three to four times per week', 'once or twice a week', 'one to three times a month', 'special occasions only', 'never', and 'prefer not to answer'. Further alcohol intake was assessed from 12 questions that participants were asked about their monthly and weekly intake (6 each) for different categories of alcohol. They included questions about red wine (data field 1568 and 4407) 'in an average week/month, how many glasses of red would you drink? (there are six glasses in an average bottle)', white wine or champagne (data fields 1578 and 4418) 'in an average week/month, how many glasses of white wine or champagne would you drink? (there are

six glasses in an average bottle)', beer plus cider (data fields 1588 and 4429) 'in an average week/month, how many pints of beer or cider would you drink? (include bitter, lager, stout, ale, Guinness)', spirits (data fields 1598 and 4440) 'in an average week/week, how many measures of spirits or liqueurs would you drink? (There are 25 standard measures in a normal sized bottle; spirits include drinks such as whisky, gin, rum, vodka brandy)' and fortified wine (data fields 1608 and 4451) 'in an average week/month how many glasses of fortified wine would you drink? (there are 12 glasses in an average bottle) (fortified wines include drinks such as sherry, port, vermouth)'. Red wine, white wine, spirits, and fortified wines were assigned an average of 10g alcohol per serving; beer, and cider 20g alcohol per serving. Total drinking per day was calculated from total alcohol intake per week and month, where weekly intakes were unavailable. Total drinking was categorised into four groups: 'never-drinkers', '<1g/day', '1-<10g/day', '10-<20g/day', '≥20g/day', and 'unknown' (0.7%). Participants that reported their overall drinking frequency was drinking only at special occasions were added to the group of those drinking less than 1g/day and those reporting never drinking alcohol were categorised as never-drinkers.

## 3.10 Summary

This chapter described the UK Biobank cohort resource, the derivation of the analytic study populations, ascertainment of cancer incidence, and the principal statistical methods used throughout this thesis. The primary analytical cohort comprised 470,771 participants, among whom 32,315 incident cancers across 28 cancer endpoints were identified during follow-up. Detailed descriptions were provided for the exposure definition of sex, participant exclusions, cancer endpoint classification, statistical power calculations, and derivation and categorisation of baseline covariates used in multivariable-adjusted analyses.

The principal statistical approach applied throughout the thesis was Cox proportional hazards regression using attained age as the underlying timescale to estimate sex-specific differences in cancer incidence. Sequentially adjusted models were used to evaluate the extent to which adjustment for sociodemographic, behavioural, anthropometric, and cancer-specific risk factors attenuated observed sex differences in cancer risk. Additional descriptive assessments of attenuation across models were conducted using percentage changes in the natural logarithm of the hazard ratio (%Δln(HR)) and, for selected analyses, Z-scores.

Together, these methodological approaches provided the analytical framework for the pan-cancer and site-specific investigations presented in Chapters 4–6. Subsequent chapters build upon this framework to examine the contribution of established risk factors, adiposity-related traits, and smoking-related exposures to observed sex differences in cancer incidence.

# Chapter 4 Sex Differences in Cancer Incidence: Prospective Analyses in the UK Biobank

## 4.1 Introduction

Most cancers common to both sexes exhibit some level of sex difference, with higher incidence and mortality generally higher in men than in women (Cook *et al.*, 2009, Edgren *et al.*, 2012, Haupt *et al.*, 2021, National Cancer Institute, 2025, National Disease Registration Service, 2025). In men, the risk is typically 1.5 to 3 times greater, except for cancers of the anus, gallbladder, and thyroid, where the incidence is higher in women (Cook *et al.*, 2009, Edgren *et al.*, 2012, Jackson *et al.*, 2022). Cancers of the bladder, oesophagus, oral cavity, and liver have nearly threefold higher incidence in men (Cook *et al.*, 2009, Edgren *et al.*, 2012, Jackson *et al.*, 2022). Between 2016 and 2018, age-standardised incidence rates for all cancers among adults aged 35–69 in the UK (excluding the two sites with the largest incidence, breast and prostate) were 363.8 per 100,000 for women and 422.3 per 100,000 for men (Shelton *et al.*, 2024). Over the last 25 years, these cancer incidence rates have initially declined and then stabilised in men, while a modest annual increase has been observed in women (Shelton *et al.*, 2024). The cancer sites with rising incidence rates in both sexes include cancers of the oral cavity, liver, kidney and melanoma, along with lung cancer in women (Shelton *et al.*, 2024), all sites which exhibit well-documented sex differences (Cook *et al.*, 2009, Edgren *et al.*, 2012, Jackson *et al.*, 2022).

The differences in cancer incidence between sexes have been mostly attributed to differences in exposure to established risk factors, which are largely anthropometric (height and obesity), health-related (smoking and alcohol consumption), molecular (chronic inflammatory and endocrinological risk factors) and genetic (EXITS genes, mosaic Y-loss)

(Adami *et al.*, 2018, Haupt *et al.*, 2021, Jackson *et al.*, 2022, Mauvais-Jarvis *et al.*, 2020). Statistics on sex differences in cancer incidence are usually reported on a national registry level, and they do not account for the hypothesised individual-level cancer risk factors (Cook *et al.*, 2009, Dong *et al.*, 2020, Edgren *et al.*, 2012, Zheng *et al.*, 2019). There have also been limited large-scale cross-cancer-site prospective cohort analyses of sex differences in cancer risk (Jackson *et al.*, 2022).

Exploring the underlying sex differences in cancer incidence and the potential contributors for the differences may provide insights into cancer aetiology and further research directions. The aim of this first research chapter was to carry out the first UK-based comprehensive prospective analyses to examine the differences in incidence of cancers (at 15 shared anatomic sites and 13 subtypes) between men and women in the UK Biobank (UKB) and to assess the extent to which any disparities persisted after accounting for established and measured risk factors for those cancers. Performing an outcome-wide analysis in one cohort allows analyses to be standardised across cancers.

The selection of cancer risk factors was guided by an *a priori* rationale informed by an existing large-scale outcome-wide analysis in the NIH–AARP Diet and Health Study (Jackson *et al.*, 2022), the *Textbook of Cancer Epidemiology* (3rd edition) (Adami *et al.*, 2018), the classifications of agents with sufficient or limited evidence of carcinogenicity as defined by the International Agency for Research on Cancer (IARC) Monographs (Volumes 1–140) (International Agency for Research on Cancer, 2025b), and the literature review on sex differences in cancer incidence undertaken in this thesis (Chapter 2). Extrinsic factors comprising sociodemographic characteristics (ethnicity and educational qualifications) and health-related behaviours (smoking and alcohol consumption), while intrinsic biological factors including body mass index (BMI) and height, were deemed important to evaluate across all analyses, reflecting their established relevance to cancer risk.

Additionally, for several cancer sites, site-specific risk factors were evaluated. Risk factors were selected based on three considerations: evidence as established or emerging risk factors, epidemiological evidence supporting their contribution to observed sex differences, and availability of relevant data within the UK Biobank cohort.

## 4.2 Methods

Pan-cancer analyses for 28 cancer endpoints—including 15 main cancer sites and 13 subtypes—in the UK Biobank cohort were selected, each with at least 100 incident cases (cases totalling in both men and women). First, a minimally-adjusted Cox model was used with stratification by age group at recruitment, Townsend index of deprivation, and region.

Second, a baseline multivariable-adjusted Cox model was used, which additionally adjusted for qualification status, ethnicity, height, smoking, and alcohol. Categorical adjustment variables were used throughout to minimise any undue influence of extreme values, except for height for which a continuous variable was used. Details of the UK Biobank cohort, exclusion criteria, exposure (sex) definition, ascertainment of cancer cases, selection of cancer endpoints for pan-cancer analyses, and details of the minimally-adjusted Cox model and the baseline multivariable-adjusted Cox model have been given in Chapter 3.

### 4.2.1 Cancer Site-specific Multivariable-adjusted Cox Models

Cancer site-specific covariates were added to the baseline multivariable-adjusted Cox models (also occasionally referred to as ‘fully adjusted models’ in the thesis) (Figure 4.1 and Table 4.1).

Cancer site-specific covariates included prevalent diabetes mellitus (for cancers of the gastric cardia, colorectum, colon, and rectum), inflammatory bowel disease and

processed meat intake (for cancers of the colorectum, colon, and rectum), prevalent HIV and number of lifetime sexual partners (for cancer of the anus), alcoholic liver disease and non-alcoholic fatty liver disease (for cancers of the liver), gallbladder disease (for cancers of the gallbladder and liver), use of UV protection, time spent outdoors in the summer, and ease of skin tanning (all for melanoma), hypertension (for cancers of the kidney), and goitre (for cancers of the thyroid). Number of sexual partners was coded as a proxy for probable exposure to human papillomavirus (HPV) infections. For the cancer-specific multivariable adjusted analyses, 51 participants were further excluded from the cohort who had a history of gastrectomy at recruitment as these individuals had a lower risk of cancers of the pancreas (associated with complete gastrectomy) and stomach (partial gastrectomy). Details on the International Classification of Diseases (ICD-10 and ICD-9) and operating procedure codes (OPCS-4) used for coding prevalent diseases and operations have been provided in Appendix Table 4A. All conditions were identified from a maximal sample analysis of 470,771 participants. Furthermore, participants also self-reported diseases during the interview portion of their assessment conducted by a medical professional at the recruitment centre on their past and current medical conditions (data-field 20002) (Appendix Table 4A). Multiple testing was corrected for using the Bonferroni method (0.05/28 tests), with $p<0.0017$ considered statistically significant.

Tests based on Schoenfeld residuals and log-log plots were used to check for violations of the proportional hazards assumption for each cancer-site specific multivariable-adjusted Cox model. There was no evidence of significant violation of the proportional hazards assumption, except for breast cancer and melanoma. Violations of the proportional hazards assumption were not explored further for breast cancer due to the rarity of this cancer in men. For melanoma, the effect of sex varied with attained age ($p<0.0001$), and so the multivariable-adjusted Cox model was rerun separately in those aged <65 years and those aged ≥65 years (threshold selected based on median age at melanoma diagnosis).

### 4.2.2 Sensitivity Analyses

Sensitivity analyses were performed on the cancer-site specific multivariable-adjusted Cox models to assess residual confounding. First, the analyses were restricted to self-reported never-smokers at recruitment (Table 4.2). Additionally, the analysis included adjusting for exposure to second hand smoke (for lung adenocarcinoma), which was measured by calculating the number of hours per week participants who identified as never smokers were exposed to tobacco smoke at home and outside home, and residential air pollution (particulate matter air pollution; PM2.5) (variable coding described in Section 3.9.8). PM2.5 was selected as the air pollution measure on the basis of its established association with lung cancer in never smokers, the availability of validated exposure estimates for the full UKB cohort, and to address residual confounding rather than as a candidate contributor for the observed sex difference (further detailed in Chapter 6).

Second, analyses were restricted to light drinkers i.e. participants who reported drinking <20 grams/day at recruitment (not including participants who reported as never drinkers); the multivariable-adjusted Cox models also included adjustment for alcohol categories for this analysis (Table 4.3). Third, analyses were restricted to participants who were of normal BMI ranges (<25 $kg/m^2$) at recruitment (Table 4.4). Fourth, analyses were restricted to participants in the highest qualification attainment group (college or university degree/vocational qualification) (Appendix Table 4B). Fifth, a quadratic term for height was additionally included, to allow for a non-linear effect (Appendix Table 4C). Sixth, analyses were additionally adjusted for the effects of physical activity, classified by low (<10 excess metabolic equivalents (METS)/week), moderate (≥10 to <50 excess METS/week), and high (≥50 excess METS/week), with a separate category for missing values (Appendix Table 4D). Seventh, fully adjusted Cox models were rerun with stratification by three broader geographic regions (England, Wales, and Scotland), rather than the ten-region classification used in the primary analyses, to assess the consistency of results by regional

stratification (Appendix Table 4E). Finally, the percentage change in the natural logarithm of the hazard ratio (%Δln(HR)) across minimally adjusted and fully adjusted models was examined to quantify the impact of covariate adjustment on effect estimates (Appendix Table 4F).

Physical activity was assessed using a list of questions within the baseline touchscreen questionnaire (data category 54) that were adapted from the short form International Physical Activity Questionnaire (IPAQ) (Craig *et al.*, 2003). The questionnaire assesses frequency, duration and of three types of activity (walking (data field 864), moderate activity (data field 884), and vigorous activity (data field 904)). In the UK Biobank cohort, the question on durations was only asked if the participants answered the question on frequency with a number other than zero. Total physical activity was then computed as the sum of walking (2.3 excess metabolic equivalents (METS)), moderate activity (3.0 excess METS), and vigorous activity (7.0 METS) (for at least 10 minutes continuously). If an activity category was missing it was treated as zero metabolic equivalents. Excess METs were categorised into low (<10 METS/week), moderate (10-<50 METS/week), and high (≥50 METS/week) groups. Excess METS represent the energy levels above the basal metabolic resting rates. These cut-offs were obtained with reference from two other papers that had classified physical activity levels in the UK Biobank (Craig *et al.*, 2003, Papier *et al.*, 2021, Perez-Cornago *et al.*, 2017). The corresponding results for these sensitivity analyses are presented in Appendix Tables 4B–4I. References to the results figures and tables have been made throughout the methods section for ease of reference.

## 4.3 Results

### 4.3.1 UK Biobank Participants

The UK Biobank study cohort included 470,771 participants aged 37 to 73 years at recruitment, 253,344 (53.8%) women and 217,427 (46.2%) men (Tables 3.1, 3.2, and 4.1). On average women were aged 56 (SD 8) and men were aged 56 (SD 8) at recruitment. During an average 10.5 (SD 2.2) years of follow up (10.6 (SD 2.1) years for women and 10.4 (SD 2.4) years for men), 32,315 incident cancers across the shared anatomical sites included in the present analysis were identified; 18,774 cancers (58.1%) in women and 13,541 (41.9%) in men. Among these cancers, the five most common cancers in men were colon, lung, melanoma, rectum, and non-Hodgkin lymphoma. The five most common cancers in women were breast, lung, colon, melanoma, and non-Hodgkin lymphoma. The analytic cohort characteristics (Table 3.1) and the cancer endpoints and case counts (Table 3.2) have been further detailed in Chapter 3 (Sections 3.2 and 3.6).

### 4.3.2 Associations Between Sex and Cancer Endpoints

Significant differences in risk by sex were observed for many cancers; men were at greater risk of cancer at most sites than women (Figure 4.1 and Table 4.1). In the minimally adjusted models, the highest HRs for cancer in men compared to women were for oesophageal adenocarcinoma (HR 6.01; 95% CI 4.90-7.37), followed by cancers of the gastric cardia (HR 4.17; 95% CI 3.15-5.53), and bladder (HR 4.07; 95% CI 3.53-4.71). The lowest HRs in men compared to women were for cancers of the breast (HR 0.007; 95% CI 0.006-0.009), followed by thyroid (HR 0.43; 95% CI 0.34-0.53), and anus (HR 0.62 95% CI 0.46-0.83).

In the multivariable models with additional adjustment for cancer site-specific risk factors, men had a higher risk than women of oesophageal adenocarcinoma (HR 5.45; 95% CI, 4.18-7.12), and of cancers of the gastric cardia (HR 3.65; 95% CI 2.48-5.38), bladder (HR 3.47;

95% CI 2.85-4.24), oral cavity (HR 2.06; 95% CI 1.69-2.51), liver (HR 1.91; 95% CI 1.48-2.47), kidney (HR 1.77; 95% CI 1.51-2.09), rectum (HR 1.70; 95% CI 1.47-1.96), and leukaemia (HR 1.43; 95% CI 1.21-1.69) (Figure 4.1 and Table 4.1). In men compared to women, the lowest risks after breast cancer were observed for cancers of the thyroid (HR 0.36; 95% CI, 0.26-0.49), anus (HR 0.41; 95% CI, 0.26-0.64), and lung adenocarcinoma (HR 0.72; 95% CI 0.62-0.84).

For ten cancer sites, the sex differences were attenuated in multivariable-adjusted models and were no longer significant at $p<0.0017$ (Figure 4.1 and Table 4.1). These were cancers of the colon (HR 1.04; 95% CI 0.94-1.15), lung (HR 0.93; 95% CI 0.84-1.03), lung small cell carcinoma (HR 1.19; 95% CI 0.86-1.64), pancreas (HR 1.04; 95% CI 0.87-1.25), Hodgkin lymphoma (HR 1.29; 95% CI 0.77-2.17), NHL (HR 1.04; 95% CI 0.91-1.19), melanoma (HR 0.97; 95% CI 0.86-1.09), gastric non-cardia (HR 1.69; 95% CI 1.09-2.60), gallbladder (HR 0.55; 95% CI 0.31-0.97), and multiple myeloma (HR 1.35; 95% CI 1.10-1.65). The multivariable-adjusted Cox model for melanoma was run separately in those aged <65 years and ≥65 years, to address violations of the proportional hazards assumption. For participants aged <65 years (including 1,177 cases), risk of cancer was lower in men compared to women (HR = 0.73, 95% CI 0.61-0.87: p-value=0.001). For those aged ≥65 years (1,529 cases), risk of cancer was nominally higher in men compared to women (HR = 1.21, 95% CI 1.03-1.42; p-value=0.01).

### 4.3.3 Sensitivity Analyses

Sex differences in cancer risk were examined in the subset of never smokers with 257,390 participants of whom 150,996 (58.7%) were women and 106,394 (41.3%) men (Table 4.2). Among them, 14,659 participants were diagnosed with a malignant cancer of interest, of whom 9,881 (67.4%) were women and 4,778 (32.6%) men. Not all endpoints had enough cases for a robust analysis; however, the HRs were mostly consistent when compared with

the main multivariable-adjusted analyses. Cancer of the stomach had a HR attenuated closer to 1 (HR 1.5; 95% CI 1.03-2.2, p=0.037), however there were not enough cancer cases to detect whether this was due to a greater proportion of gastric non-cardia cancers in never smokers, or some residual effects of smoking in the multivariable-adjusted models.

Second, sex differences in cancer risk were examined in the subset of light alcohol drinkers with 301,509 participants, of whom 192,999 (64.0%) were women and 108,510 (36.0%) men (Table 4.3). Among them, 20,286 participants were diagnosed with a malignant cancer of interest, of whom 14,037 (69.2%) were women and 6,249 (30.8%) men. After multivariable adjustments, the HRs were consistent with those in the main analyses for most cancer endpoints, except for cancers of the oral cavity (HR 1.54; 95% CI 1.17-2.02, p=0.002), which somewhat attenuated.

Third, sex differences in cancer risk were examined in the subset of 152,846 participants who were of normal BMI (<25kg/$m^2$) at baseline, of whom 99,284 (65.0%) were women and 53,579 (35.0%) were men (Table 4.4). Among them, 9,956 participants were diagnosed with a malignant cancer of interest, of whom 6,992 (70.2%) were in women and 2964 (29.8%) were in men. After multivariate adjustments, the HRs were consistent with those in the main analyses for most cancer endpoints, except for cancer of the liver (HR 1.18; 95% CI 0.70-1.97, p=0.54), which attenuated to the null. For rectal cancer, the association was somewhat attenuated but remained statistically significant at Bonferroni corrected p<0.0017 (HR 1.54; 95% CI 1.18–2.01, p=0.001).

Fourth, sex differences in cancer risk were examined in the subset of 279,479 participants with the highest qualification attainment group (Appendix Table 4B), of whom 143,373 (51.3%) were women and 136,106 (48.7%) were men. After multivariate adjustments, the HRs were consistent with those in the main analyses for most cancer endpoints, except for

leukaemia (HR 1.33; 95% CI 1.06-1.68, p=0.015), for which the association was attenuated and was no longer statistically significant after Bonferroni correction.

Additional sensitivity analyses indicated that further adjustment for height using a quadratic term, as well as adjustment for physical activity, did not materially alter the hazard ratios for any cancer endpoint (Appendix Tables 4C and 4D). Similarly, re-fitting the multivariable cancer site-specific Cox models with stratification by broader geographic regions (England, Wales, and Scotland), rather than the ten-region classification used in the primary analyses, also did not materially alter the hazard ratios for any cancer endpoint (Appendix Table 4E). The percentage change in the natural logarithm of HRs (ln(HR)) between minimally adjusted and site-specific multivariable models was examined (Appendix Table 4F). For cancer sites with marked sex differences that remained consistent across sensitivity analyses, attenuation was modest for oesophageal adenocarcinoma (5.5%) and bladder cancer (11.4%), and higher for kidney (23.5%) and rectal cancer (27.5%). For lung adenocarcinoma, percent attenuation was not calculated as the minimally adjusted estimate was close to the null; instead, the direction and magnitude were interpreted qualitatively. The attenuation was linked to smoking, as adjustment for smoking revealed the higher risk for women.

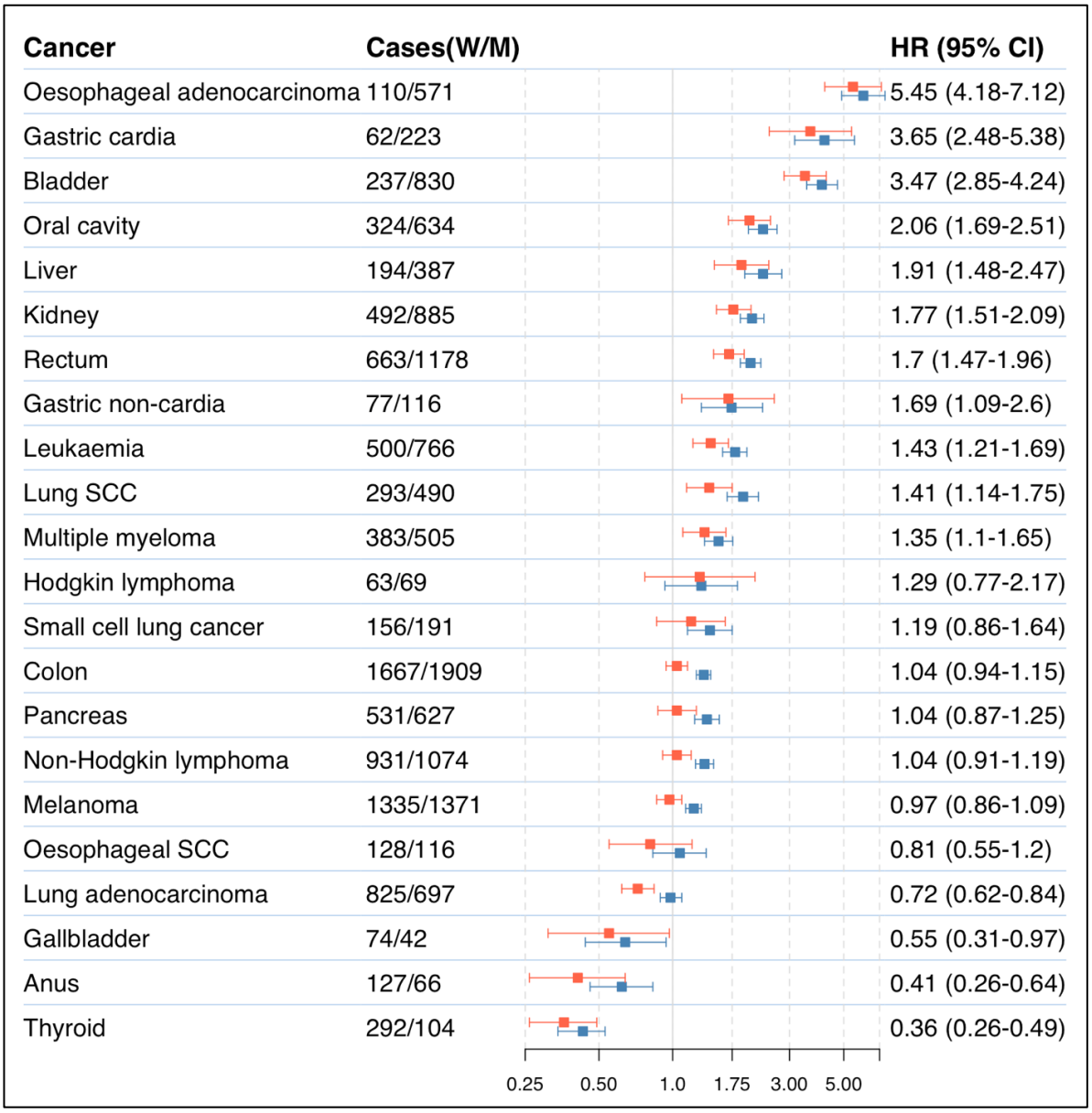



**Figure 4.1 Hazard ratios (HR) and 95% confidence intervals (CI) for incident cancer in men compared to women in UK Biobank.**

**Minimally-adjusted (blue) and cancer-specific multivariable-adjusted Cox models (orange)*. M men; SCC squamous cell carcinoma; W women.**

**Table 4.1 Hazard ratios (HRs) and 95% confidence intervals (CIs) comparing risk of cancer in men to women.**

| | | | | Minimally-adjusted model[a] | | Multivariable-adjusted model[b] | | Multivariable-adjusted model with cancer specific covariates[c] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Women (%) | Men (%) | Total count | HR (95% CI) | p-value | HR (95% CI) | p-value | Additional covariates | HR (95% CI) | p-value |
| **Oral cavity [C00-C14]** | 324 (33.8) | 634 (66.2) | 958 | 2.34 (2.04-2.67) | <0.0001 | 2.06 (1.69-2.51) | <0.0001 | | 2.06 (1.69-2.51) | <0.0001 |
| **Oesophagus [C15]** | 249 (26.0) | 707 (74.0) | 956 | 3.3 (2.86-3.82) | <0.0001 | 2.89 (2.36-3.55) | <0.0001 | | 2.89 (2.36-3.55) | <0.0001 |
| **Adenocarcinoma** | 110 (16.2) | 571 (83.8) | 681 | 6.01 (4.9-7.37) | <0.0001 | 5.47 (4.19-7.14) | <0.0001 | Gastro-oesophageal reflux disease (GORD) | 5.45 (4.18-7.12) | <0.0001 |
| **Squamous cell carcinoma** | 128 (52.5) | 116 (47.5) | 244 | 1.07 (0.83-1.37) | 0.622 | 0.81 (0.55-1.2) | 0.297 | | 0.81 (0.55-1.2) | 0.297 |
| **Stomach [C16]** | 203 (31.3) | 446 (68.7) | 649 | 2.55 (2.16-3.01) | <0.0001 | 2.34 (1.83-2.97) | <0.0001 | | 2.34 (1.83-2.97) | <0.0001 |
| **Gastric cardia [C16.0]** | 62 (21.8) | 223 (78.2) | 285 | 4.17 (3.15-5.53) | <0.0001 | 3.7 (2.51-5.44) | <0.0001 | Diabetes | 3.65 (2.48-5.38) | <0.0001 |
| **Gastric non-cardia [C16.1-C16.6]** | 77 (39.9) | 116 (60.1) | 193 | 1.74 (1.31-2.33) | <0.0001 | 1.69 (1.09-2.6) | 0.018 | | 1.69 (1.09-2.6) | 0.018 |
| **Colorectum [C18-C20]** | 2316 (43.0) | 3072 (57.0) | 5388 | 1.55 (1.47-1.64) | <0.0001 | 1.25 (1.15-1.36) | <0.0001 | Diabetes, inflammatory bowel disease, processed meat | 1.22 (1.12-1.33) | <0.0001 |

| | | | | Minimally-adjusted model[a] | | Multivariable-adjusted model[b] | | Multivariable-adjusted model with cancer specific covariates[c] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Women (%) | Men (%) | Total count | HR (95% CI) | p-value | HR (95% CI) | p-value | Additional covariates | HR (95% CI) | p-value |
| **Colon [C18]** | 1667 (46.6) | 1909 (53.4) | 3576 | 1.34 (1.25-1.43) | <0.0001 | 1.06 (0.96-1.17) | 0.254 | Diabetes, inflammatory bowel disease, processed meat | 1.04 (0.94-1.15) | 0.456 |
| **Rectum [C19-C20]** | 663 (36.0) | 1178 (64.0) | 1841 | 2.08 (1.89-2.29) | <0.0001 | 1.74 (1.51-2.01) | <0.0001 | Diabetes, inflammatory bowel disease, processed meat | 1.7 (1.47-1.96) | <0.0001 |
| **Anus [C21]** | 127 (65.8) | 66 (34.2) | 193 | 0.62 (0.46-0.83) | 0.001 | 0.45 (0.29-0.71) | <0.0001 | Human immunodeficiency virus (HIV), no. of sexual partners | 0.41 (0.26-0.64) | <0.0001 |
| **Liver [C22]** | 194 (33.4) | 387 (66.6) | 581 | 2.34 (1.97-2.79) | <0.0001 | 2 (1.55-2.58) | <0.0001 | Alcoholic liver disease, non-alcoholic fatty liver disease, gallbladder disease | 1.91 (1.48-2.47) | <0.0001 |
| **Gallbladder [C23]** | 74 (63.8) | 42 (36.2) | 116 | 0.64 (0.44-0.94) | 0.023 | 0.56 (0.31-0.99) | 0.046 | Gallbladder disease | 0.55 (0.31-0.97) | 0.0036 |
| **Pancreas [C25]** | 531 (45.9) | 627 (54.1) | 1158 | 1.38 (1.23-1.55) | <0.0001 | 1.04 (0.87-1.25) | 0.642 | | 1.04 (0.87-1.25) | 0.642 |
| **Lung [C34]** | 1779 (48.6) | 1885 (51.4) | 3664 | 1.23 (1.16-1.32) | <0.0001 | 0.93 (0.84-1.03) | 0.144 | | 0.93 (0.84-1.03) | 0.144 |

| | | | | Minimally-adjusted model[a] | | Multivariable-adjusted model[b] | | Multivariable-adjusted model with cancer specific covariates[c] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Women (%) | Men (%) | Total count | HR (95% CI) | p-value | HR (95% CI) | p-value | Additional covariates | HR (95% CI) | p-value |
| **Adenocarcinoma** | 825 (54.2) | 697 (45.8) | 1522 | 0.98 (0.89-1.09) | 0.759 | 0.72 (0.62-0.84) | <0.0001 | | 0.72 (0.62-0.84) | <0.0001 |
| **Squamous cell carcinoma** | 293 (37.4) | 490 (62.6) | 783 | 1.94 (1.67-2.24) | <0.0001 | 1.41 (1.14-1.75) | 0.002 | | 1.41 (1.14-1.75) | 0.002 |
| **Small cell carcinoma** | 156 (45.0) | 191 (55.0) | 347 | 1.42 (1.15-1.75) | 0.001 | 1.19 (0.86-1.64) | 0.289 | | 1.19 (0.86-1.64) | 0.289 |
| **Melanoma [C43]** | 1335 (49.3) | 1371 (50.7) | 2706 | 1.22 (1.13-1.31) | <0.0001 | 0.87 (0.77-0.98) | 0.019 | Use of UV protection, time spent outdoors in the summer, ease of skin tanning | 0.97 (0.86-1.09) | 0.591 |
| **Breast [C50]** | 8737 (99.4) | 55 (0.6) | 8792 | 0.007 (0.006-0.009) | <0.0001 | 0.005 (0.004-0.007) | <0.0001 | | 0.005 (0.004-0.007) | <0.0001 |
| **Kidney [C64-C65]** | 492 (35.7) | 885 (64.3) | 1377 | 2.11 (1.89-2.36) | <0.0001 | 1.84 (1.56-2.16) | <0.0001 | Hypertension | 1.77 (1.51-2.09) | <0.0001 |
| **Bladder [C67]** | 237 (22.2) | 830 (77.8) | 1067 | 4.07 (3.53-4.71) | <0.0001 | 3.47 (2.85-4.24) | <0.0001 | | 3.47 (2.85-4.24) | <0.0001 |
| **Thyroid [C73]** | 292 (73.7) | 104 (26.3) | 396 | 0.43 (0.34-0.53) | <0.0001 | 0.35 (0.25-0.48) | <0.0001 | Goitre | 0.36 (0.26-0.49) | <0.0001 |

| | | | | Minimally-adjusted model[a] | | Multivariable-adjusted model[b] | | Multivariable-adjusted model with cancer specific covariates[c] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Women (%) | Men (%) | Total count | HR (95% CI) | p-value | HR (95% CI) | p-value | Additional covariates | HR (95% CI) | p-value |
| **Lymphatic and haematopoietic cancers [C81-C96]** | 1886 (43.7) | 2431 (56.3) | 4317 | 1.51 (1.42-1.60) | <0.0001 | 1.22 (1.11-1.33) | <0.0001 | | 1.22 (1.11-1.33) | <0.0001 |
| **Hodgkin lymphoma [C81]** | 63 (47.7) | 69 (52.3) | 132 | 1.31 (0.93-1.84) | 0.122 | 1.29 (0.77-2.17) | 0.326 | | 1.29 (0.77-2.17) | 0.326 |
| **Non-Hodgkin lymphoma [C82-C85]** | 931 (46.4) | 1074 (53.6) | 2005 | 1.35 (1.24-1.47) | <0.0001 | 1.04 (0.91-1.19) | 0.585 | | 1.04 (0.91-1.19) | 0.585 |
| **Multiple myeloma [C90 & C88]** | 383 (43.1) | 505 (56.9) | 888 | 1.54 (1.35-1.76) | <0.0001 | 1.35 (1.1-1.65) | 0.004 | | 1.35 (1.10-1.65) | 0.004 |
| **Leukaemia [C91-C95]** | 500 (39.5) | 766 (60.5) | 1266 | 1.8 (1.60-2.01) | <0.0001 | 1.43 (1.21-1.69) | <0.0001 | | 1.43 (1.21-1.69) | <0.0001 |

[a]Minimally-adjusted hazard ratios (95% CIs) stratifying for age, fifths of Townsend deprivation index and region.

[b]Multivariable-adjusted hazard ratios (95% CIs), additionally adjusted for ethnicity, qualifications, height, BMI, smoking (status and intensity), and alcohol.

[c]Multivariable-adjusted hazard ratios (95% CIs) with additional adjustment for cancer specific covariates. Covariate selection has been described in the text.

*p-value was not retrievable further. P-values <0.001 are displayed as “p<0.001” for readability. Appendix Table 4J reproduces Table 4.1 with exact p-values.

**Table 4.2 Cancer specific multivariable-adjusted Cox proportional hazards models comparing risk of cancer incidence in men versus women who identified as never smokers at recruitment in the UK Biobank.**

| | Women (%) | Men (%) | Total | HR (95% CI) | P value | Main analysis results* |
|---|---|---|---|---|---|---|
| **Oral cavity [C00-C14]** | 138 (42.6) | 186 (57.4) | 324 | 1.79 (1.29-2.50) | 5.57e-04 | 2.06 (1.69-2.51) |
| **Oesophagus [C15]** | 117 (37.7) | 193 (62.3) | 310 | 2.61 (1.85-3.68) | 4.87e-08 | 2.89 (2.36-3.55) |
| **Adenocarcinoma** | 48 (23.6) | 155 (76.4) | 203 | 5.11 (3.25-8.02) | 1.49e-12 | 5.45 (4.18-7.12) |
| **Squamous cell carcinoma** | 66 (71.7) | 26 (28.3) | 92 | *[insufficient data]* | *[insufficient data]* | 0.81 (0.55-1.2) |
| **Stomach [C16]** | 110 (44) | 140 (56) | 250 | 1.5 (1.03-2.2) | 0.037 | 2.34 (1.83-2.97) |
| **Gastric cardia [C16.0]** | 25 (29.4) | 60 (70.6) | 85 | *[insufficient data]* | *[insufficient data]* | 3.65 (2.48-5.38) |
| **Gastric non-cardia [C16.1-C16.6]** | 45 (53.6) | 39 (46.4) | 84 | *[insufficient data]* | *[insufficient data]* | 1.69 (1.09-2.6) |
| **Colorectum [C18-C20]** | 1266 (50.5) | 1242 (49.5) | 2508 | 1.24 (1.1-1.4) | 3.862e-25 | 1.22 (1.12-1.33) |
| **Colon [C18]** | 936 (54.9) | 768 (45.1) | 1704 | 1.01 (0.87-1.17) | 0.875 | 1.04 (0.94-1.15) |
| **Rectum [C19-C20]** | 337 (41.5) | 476 (58.5) | 813 | 1.91 (1.54-2.36) | 2.78e-09 | 1.7 (1.47-1.96) |
| **Anus [C21]** | 55 (70.5) | 23 (29.5) | 78 | *[insufficient data]* | *[insufficient data]* | 0.41 (0.26-0.64) |
| **Liver [C22]** | 92 (43.6) | 119 (56.4) | 211 | 2.39 (1.58-3.62) | 3.67e-05 | 1.91 (1.48-2.47) |
| **Gallbladder [C23]** | 31 (68.9) | 14 (31.1) | 45 | *[insufficient data]* | *[insufficient data]* | 0.55 (0.31-0.97) |
| **Pancreas [C25]** | 274 (52.6) | 247 (47.4) | 521 | 1.13 (0.87-1.47) | 0.369 | 1.04 (0.87-1.25) |
| **Lung [C34]** | 316 (65.7) | 165 (34.3) | 481 | 0.71 (0.53-0.94) | 0.017 | 0.93 (0.84-1.03) |
| **Adenocarcinoma** | 174 (68.5) | 80 (31.5) | 254 | 0.62 (0.42-0.93) | 0.002 | 0.72 (0.62-0.84) |
| **Squamous cell carcinoma** | 31 (64.6) | 17 (35.4) | 48 | *[insufficient data]* | *[insufficient data]* | 1.41 (1.14-1.75) |
| **Small cell carcinoma** | 5 (62.5) | 3 (37.5) | 8 | *[insufficient data]* | *[insufficient data]* | 1.19 (0.86-1.64) |
| **Melanoma [C43]** | 823 (53.7) | 710 (46.3) | 1533 | 0.88 (0.75-1.03) | 0.103 | 0.97 (0.86-1.09) |

| | Women (%) | Men (%) | Total | HR (95% CI) | P value | Main analysis results* |
|---|---|---|---|---|---|---|
| **Breast [C50]** | 5029 (99.5) | 24 (0.5) | 5053 | 0.005 (0.003-0.007) | <0.0e-310 | 0.005(0.004-0.007) |
| **Kidney [C64-65]** | 252 (42.7) | 338 (57.3) | 590 | 2.10 (1.64-2.69) | 3.65e-09 | 1.77 (1.51-2.09) |
| **Bladder [C67]** | 104 (31.1) | 230 (68.9) | 334 | 3.15 (2.25-4.41) | 1.98e-11 | 3.47 (2.85-4.24) |
| **Thyroid [C73]** | 191 (79.6) | 49 (20.4) | 240 | 0.35 (0.23-0.54) | 2.08e-06 | 0.36 (0.26-0.49) |
| **Lymphatic and haematopoietic [C81-C96]** | 1083 (49.7) | 1098 (50.3) | 2181 | 1.15 (1.01-1.31) | 0.029 | 1.22 (1.11-1.33) |
| **Hodgkin lymphoma [C81]** | 34 (52.3) | 31 (47.7) | 65 | *[insufficient data]* | *[insufficient data]* | 1.29 (0.77-2.17) |
| **Non-Hodgkin lymphoma [C82-C85]** | 547 (52.7) | 491 (47.3) | 1038 | 0.92 (0.76-1.11) | 0.387 | 1.04 (0.91-1.19) |
| **Multiple myeloma [C90 & C88]** | 221 (47) | 249 (53) | 470 | 1.36 (1.03-1.79) | 0.028 | 1.35 (1.10-1.65) |
| **Leukaemia [C91-C95]** | 274 (46.2) | 319 (53.8) | 593 | 1.44 (1.13-1.85) | 0.003 | 1.43 (1.21-1.69) |

This sensitivity analysis was performed on a multivariable-adjusted model that included cancer-specific, and baseline set of covariates. Covariate selection has been described in text and supplementary tables. Cancer endpoints were excluded from the analysis if total counts were less 100. **257,390** (58.7% women) participants self-reported as never smokers, of which 14,659 were diagnosed with an incident malignant cancer of interest, with 9,881 cases (67.4%) in women.

*The main analysis results (Table 4.1) for all participants (HRs and 95% CIs from multivariable-adjusted models with cancer specific covariates) are also shown for comparison.

**Table 4.3 Cancer specific multivariable-adjusted Cox proportional hazards models comparing risk of cancer incidence in men versus women who identified as light alcohol drinkers (<20 g/day) in the UK Biobank.**

| | Women (%) | Men (%) | Total | HR (95% CI) | P value | Main analysis results* |
|---|---|---|---|---|---|---|
| **Oral cavity [C00-C14]** | 236 (51.3) | 224 (48.7) | 460 | 1.54 (1.17-2.02) | 0.002 | 2.06 (1.69-2.51) |
| **Oesophagus [C15]** | 164 (34.9) | 306 (65.1) | 470 | 2.89 (2.2-3.8) | 3.90e-14 | 2.89 (2.36-3.55) |
| **Adenocarcinoma** | 81 (23.1) | 269 (76.9) | 350 | 5.02 (3.58-7.05) | 8.78e-21 | 5.45 (4.18-7.12) |
| **Squamous cell carcinoma** | 75 (68.8) | 34 (31.2) | 109 | 0.65 (0.36-1.17) | 0.151 | 0.81 (0.55-1.2) |
| **Stomach [C16]** | 148 (42.8) | 198 (57.2) | 346 | 2.30 (1.68-3.15) | 2.23e-07 | 2.34 (1.83-2.97) |
| **Gastric cardia [C16.0]** | 45 (31.9) | 96 (68.1) | 141 | 4.01 (2.41-6.66) | 8.56e-08 | 3.65 (2.48-5.38) |
| **Gastric non-cardia [C16.1-C16.6]** | 55 (50.5) | 54 (49.5) | 109 | 1.66 (0.95-2.91) | 0.074 | 1.69 (1.09-2.6) |
| **Colorectum [C18-C20]** | 1713 (55.8) | 1355 (44.2) | 3068 | 1.2 (1.07-1.33) | 0.001 | 1.22 (1.12-1.33) |
| **Colon [C18]** | 1242 (59) | 863 (41) | 2105 | 1.02 (0.9-1.16) | 0.757 | 1.04 (0.94-1.15) |
| **Rectum [C19/C20]** | 482 (49.4) | 494 (50.6) | 976 | 1.65 (1.37-2) | 2.09e-07 | 1.7 (1.47-1.96) |
| **Anus [C21]** | 100 (79.4) | 26 (20.6) | 126 | 0.29 (0.16-0.52) | 3.96e-05 | 0.41 (0.26-0.64) |
| **Liver [C22]** | 146 (49.5) | 149 (50.5) | 295 | 1.82 (1.29-2.56) | 0.001 | 1.91 (1.48-2.47) |
| **Gallbladder [C23]** | 58 (75.3) | 19 (24.7) | 77 | *[insufficient data]* | *[insufficient data]* | 0.55 (0.31-0.97) |
| **Pancreas [C25]** | 384 (58.3) | 275 (41.7) | 659 | 0.98 (0.78-1.23) | 0.852 | 1.04 (0.87-1.25) |
| **Lung [C34]** | 1190 (60.7) | 769 (39.3) | 1959 | 0.95 (0.83-1.09) | 0.45 | 0.93 (0.84-1.03) |
| **Adenocarcinoma [C34]** | 559 (65.9) | 289 (34.1) | 848 | 0.72 (0.58-0.88) | 0.002 | 0.72 (0.62-0.84) |
| **Squamous cell carcinoma [C34]** | 188 (49.7) | 190 (50.3) | 378 | 1.45 (1.07-1.96) | 0.015 | 1.41 (1.14-1.75) |
| **Small cell carcinoma [C34]** | 102 (58.3) | 73 (41.7) | 175 | 1.24 (0.79-1.93) | 0.351 | 1.19 (0.86-1.64) |

| | Women (%) | Men (%) | Total | HR (95% CI) | P value | Main analysis results* |
|---|---|---|---|---|---|---|
| **Melanoma [C43]** | 1041 (59.7) | 703 (40.3) | 1744 | 0.97 (0.84-1.12) | 0.651 | 0.97 (0.86-1.09) |
| **Breast [C50]** | 6603 (99.6) | 28 (0.4) | 6631 | 0.006 (0.004-0.008) | 1.40e-159 | 0.005 (0.004-0.007) |
| **Kidney [C64-65]** | 396 (45.3) | 479 (54.7) | 875 | 1.67 (1.36-2.03) | 5.43e-07 | 1.77 (1.51-2.09) |
| **Bladder [C67]** | 180 (30.6) | 409 (69.4) | 589 | 3.61 (2.82-4.64) | 4.65e-24 | 3.47 (2.85-4.24) |
| **Thyroid [C73]** | 212 (78.5) | 58 (21.5) | 270 | 0.36 (0.24-0.53) | 3.65e-07 | 0.36 (0.26-0.49) |
| **Lymphatic and haematopoietic [C81-C96]** | 1466 (54) | 1251 (46) | 2717 | 1.2 (1.07-1.34) | 0.002 | 1.22 (1.11-1.33) |
| **Hodgkin lymphoma [C81]** | 47 (54.7) | 39 (45.3) | 86 | *[insufficient data]* | *[insufficient data]* | 1.29 (0.77-2.17) |
| **Non-Hodgkin lymphoma [C82-C85]** | 724 (56.8) | 550 (43.2) | 1274 | 1 (0.85-1.18) | 0.967 | 1.04 (0.91-1.19) |
| **Multiple myeloma [C90 & C88]** | 294 (52.3) | 268 (47.7) | 562 | 1.39 (1.09-1.79) | 0.008 | 1.35 (1.10-1.65) |
| **Leukaemia [C91-C95]** | 395 (50.4) | 389 (49.6) | 784 | 1.4 (1.14-1.73) | 0.002 | 1.43 (1.21-1.69) |

This sensitivity analysis was performed on a multivariable-adjusted model that included cancer-specific, and baseline set of covariates. Covariate selection has been described in text and supplementary tables. Cancer endpoints were excluded from the analysis if total counts were less 100. **301,509** participants (64.0% women) self-reported as light drinkers in the cohort, meaning they consumed <1g/d to <20g/d of alcohol (grams/day). Of these 20,286 participants had malignant cancers, with 14,037 cases (69.2%) in women.

*The main analyses results for all participants (HRs and 95% CIs from multivariable-adjusted models with cancer specific covariates) are also shown for comparison.

**Table 4.4 Cancer specific multivariable-adjusted Cox proportional hazards models comparing risk of cancer incidence in men versus women had normal BMI at baseline (<25 kg/m2) in the UK Biobank.**

| | Women (%) | Men (%) | Total count | HR (95% CI) | P value | Main analysis results* |
|---|---|---|---|---|---|---|
| **Lip, oral cavity, and pharynx [C00-C14]** | 134 (40.7) | 195 (59.3) | 329 | 2.38 (1.72-3.29) | 1.39e-07 | 2.06 (1.69-2.51) |
| **Oesophagus [C15]** | 93 (46.3) | 108 (53.7) | 201 | 1.55 (1.02-2.34) | 0.039 | 2.89 (2.36-3.55) |
| **Adenocarcinoma** | 26 (24.5) | 80 (75.5) | 106 | 4.71 (2.55-8.69) | 7.02e-07 | 5.45 (4.18-7.12) |
| **Squamous cell carcinoma** | 61 (68.5) | 28 (31.5) | 89 | *[insufficient data]* | *[insufficient data]* | 0.81 (0.55-1.2) |
| **Stomach [C16]** | 67 (46.2) | 78 (53.8) | 145 | 2.33 (1.44-3.77) | 5.95e-04 | 2.34 (1.83-2.97) |
| **Gastric cardia [C16.0]** | 17 (31.5) | 37 (68.5) | 54 | *[insufficient data]* | *[insufficient data]* | 3.65 (2.48-5.38) |
| **Gastric non-cardia [C16.1-C16.6]** | 26 (46.4) | 30 (53.6) | 56 | *[insufficient data]* | *[insufficient data]* | 1.69 (1.09-2.6) |
| **Colorectum [C18-C20]** | 826 (58) | 599 (42) | 1425 | 1.12 (0.96-1.31) | 0.161 | 1.22 (1.12-1.33) |
| **Colon [C18]** | 576 (62.3) | 349 (37.7) | 925 | 0.93 (0.76-1.14) | 0.485 | 1.04 (0.94-1.15) |
| **Rectum [C19/C20]** | 255 (50.3) | 252 (49.7) | 507 | 1.54 (1.18-2.01) | 0.001 | 1.7 (1.47-1.96) |
| **Anus [C21]** | 48 (70.6) | 20 (29.4) | 68 | *[insufficient data]* | *[insufficient data]* | 0.41 (0.26-0.64) |
| **Liver [C22]** | 68 (52.3) | 62 (47.7) | 130 | 1.18 (0.7-1.97) | 0.54 | 1.91 (1.48-2.47) |
| **Gallbladder [C23]** | 11 (61.1) | 7 (38.9) | 18 | *[insufficient data]* | *[insufficient data]* | 0.55 (0.31-0.97) |
| **Pancreas [C25]** | 165 (57.5) | 122 (42.5) | 287 | 1.04 (0.73-1.49) | 0.808 | 1.04 (0.87-1.25) |
| **Lung [C34]** | 641 (55.5) | 514 (44.5) | 1155 | 1 (0.84-1.19) | 0.984 | 0.93 (0.84-1.03) |
| **Adenocarcinoma**** | 309 (60.4) | 203 (39.6) | 512 | 0.89 (0.69-1.16) | 0.397 | 0.72 (0.62-0.84) |
| **Squamous cell carcinoma** | 105 (43.9) | 134 (56.1) | 239 | 1.39 (0.95-2.05) | 0.092 | 1.41 (1.14-1.75) |
| **Small cell carcinoma** | 41 (47.1) | 46 (52.9) | 87 | *[insufficient data]* | *[insufficient data]* | 1.19 (0.86-1.64) |

| | Women (%) | Men (%) | Total count | HR (95% CI) | P value | Main analysis results* |
|---|---|---|---|---|---|---|
| **Melanoma [C43]** | 526 (66) | 271 (34) | 797 | 0.78 (0.63-0.98) | 0.031 | 0.97 (0.86-1.09) |
| **Breast [C50]** | 3145 (99.6) | 14 (0.4) | 3159 | 0.006 (0.004-0.01) | 1.55e-79 | 0.005 (0.004-0.007) |
| **Kidney [C64-65] (WHO)** | 128 (45.7) | 152 (54.3) | 280 | 1.99 (1.40-2.83) | 1.23e-04 | 1.77 (1.51-2.09) |
| **Bladder [C67]** | 75 (29.5) | 179 (70.5) | 254 | 4.36 (2.98-6.38) | 3.06e-14 | 3.47 (2.85-4.24) |
| **Thyroid [C73]** | 93 (79.5) | 24 (20.5) | 117 | 0.35 (0.19-0.64) | 6.44e-04 | 0.36 (0.26-0.49) |
| **Lymphatic and haematopoietic [C81/C96]** | 689 (55.4) | 554 (44.6) | 1243 | 1.13 (0.96-1.34) | 0.143 | 1.22 (1.11-1.33) |
| **Hodgkin lymphoma [C81]** | 26 (63.4) | 15 (36.6) | 41 | *[insufficient data]* | *[insufficient data]* | 1.29 (0.77-2.17) |
| **Non-Hodgkin lymphoma [C82/85]** | 358 (59.5) | 244 (40.5) | 602 | 0.92 (0.72-1.17) | 0.476 | 1.04 (0.91-1.19) |
| **Multiple myeloma [C90 & C88]** | 128 (53.8) | 110 (46.2) | 238 | 1.37 (0.94-2.01) | 0.103 | 1.35 (1.10-1.65) |
| **Leukaemia [C91/C95]** | 171 (48.9) | 179 (51.1) | 350 | 1.54 (1.12-2.1) | 0.007 | 1.43 (1.21-1.69) |

This sensitivity analysis was performed on a multivariable-adjusted model that included cancer-specific, and baseline set of covariates. Covariate selection has been described in text. Cancer endpoints were excluded from the analysis if total counts were less 100. **152,846** participants (65% women) had normal BMI at baseline (<25 kg/m$^2$). Of these, 9,956 participants were diagnosed with malignant cancers, with 6,992 cases (70.2%) in women.

*The main analyses results for all participants (HRs and 95% CIs from multivariable-adjusted models with cancer specific covariates) are also shown for comparison.

**The attenuation towards the null for lung adenocarcinoma may reflect differences in smoking patterns within this restricted cohort

## 4.4 Discussion

In this first UK based comprehensive prospective analysis of sex differences in the risk of cancer performed in the UK, there were marked sex differences in the incidence of several cancers that were not explained by differences in established risk factors. Men compared to women had a higher risk of eight cancers, the largest difference being for oesophageal adenocarcinoma, followed by cancers of the gastric cardia, bladder, oral cavity, liver, kidney, and rectum, and leukaemia. Women compared to men had a higher risk of four cancers, with the highest relative risk for breast cancer, followed by cancers of the thyroid, and anus, and lung adenocarcinoma. Furthermore, for melanoma the differences in incidence by sex varied by age; under the ages of 65 women had a higher risk compared to men, whereas over the age of 65 men had a higher risk than women. For all the above-mentioned sites, the differences in risk by sex were attenuated a little but remained after further adjustment for established cancer risk factors that may play a role in contributing to the differences.

The findings of these analyses are consistent with the one previous pan-cancer epidemiological study that has investigated sex differences within a prospective cohort study (Jackson *et al.*, 2022) and with studies that have used national level registries (Cook *et al.*, 2009, Dong *et al.*, 2020, Edgren *et al.*, 2012, Radkiewicz *et al.*, 2017, Zheng *et al.*, 2019).

Explanations for sex differences in cancer incidence may include sex differences in prevalence and/or effects of other risk factors and emerging risk factors that may include dietary, molecular, genetic, and environmental factors. Previous research on sex differences in cancer incidence has suggested height (possibly as a proxy for cell number) may explain part of the sex differences (Fu *et al.*, 2020b, Walter *et al.*, 2013) and while our analyses reiterated the association of height with cancer risk, accounting for the height

differences explained some but not all the sex differences. Moreover, the results from sensitivity analyses restricted to participants who had never smoked and were light drinkers of alcohol suggested that residual confounding by smoking and alcohol, respectively, was unlikely to explain the sex differences in cancer incidence for sites including bladder, kidney, lung adenocarcinoma, oesophageal adenocarcinoma, rectum, and thyroid. Here, those cancer endpoints are discussed further along with other potential risk factors that may contribute to their respective sex differences in incidence and highlight where future research may be valuable.

### 4.4.1 Oesophageal adenocarcinoma

The excess risk of oesophageal adenocarcinoma in men, which was more than fivefold greater than women in the current study, has been attributed to a higher prevalence of obesity in men, and of gastro-oesophageal reflux disease, and their relationship with each other (Nadaleto *et al.*, 2016, Tian *et al.*, 2020). Both factors were adjusted for and neither risk factor explained much of the sex difference. However, since BMI does not capture fat distribution entirely and there are substantial sex differences in fat distribution, other potential explanations may include the role of higher abdominal obesity in men which contributes to increased chronic gastric pressure and increased cancer risk (Abnet *et al.*, 2018, Chandanos and Lagergren, 2009, Nadaleto *et al.*, 2016, Rask-Andersen *et al.*, 2023). The role of sex hormones has also been considered in relation to the sex differences; however, the epidemiological evidence is variable (reviewed further in section 2.3.4).

### 4.4.2 Bladder

The high incidence of bladder cancer in men compared to women (more than threefold in the UK Biobank) is not fully understood but may be due to a combination of differences in carcinogenic exposures (through occupations), sex hormones, and genetic differences by sex (Kogevinas *et al.*, 2018, Theodorescu *et al.*, 2022). Men are more likely to have higher

carcinogenic exposures through occupational exposures, including but not limited to occupations related to aluminium production, rubber production, painting, and firefighting (Brown *et al.*, 2018, Hadkhale *et al.*, 2019, International Agency for Research on Cancer, 2025b, Jubber *et al.*, 2023, Kogevinas *et al.*, 2018). Data on hormones suggest that oestrogens may be protective for bladder cancer in women, but the role of androgens is unclear (Daugherty *et al.*, 2013, Jubber *et al.*, 2023, Mcgrath *et al.*, 2006). Furthermore, genetic studies suggest the mosaic loss of Y chromosome in blood in men may be associated with increased bladder cancer risk in men and some X chromosome linked genes (*KDM6A*) may confer protection in women (Dunford *et al.*, 2017, Forsberg *et al.*, 2014). Men and women generally have an equal expression of X-linked genes, as during embryogenesis one of the two X chromosomes in women randomly undergoes inactivation and there is haploid expression of the X chromosome in each cell (Dunford *et al.*, 2017). However, some genes (termed EXITS genes, derived from 'escape from X-inactivation tumour suppressors') may escape the inactivation process rendering them to have biallelic expression in women. These group of genes include tumour suppressor genes (including *KDM6A*) and thus may confer protection in women as inactivation would require two deleterious mutations whereas for men it would require one (Dunford *et al.*, 2017). It has also been hypothesized that bladder cancer symptoms in women also may be mistaken for more common conditions such as urinary tract infections, for which they may be treated with several antibiotic courses first before undergoing urological investigations (Theodorescu *et al.*, 2022). However, these delays may only have a modest impact (Lyratzopoulos *et al.*, 2013) and it is unlikely that they contribute materially to the observed sex differences in incidence.

### 4.4.3 Kidney

Similar to the results of our analysis, a previous cross-sectional study (Scelo *et al.*, 2018) reported the relative risk of kidney cancer was twice as high in men compared to women, a difference that was consistent by age, year of diagnosis, and region worldwide. They hypothesized some of this sex difference may be attributed to sex differences in smoking, BMI, and hypertension, which our analyses accounted for and yet we still observed a 1.7-fold increase in risk of kidney cancer in men. Other potential risk factors may include occupational hazards including but not limited to working with herbicides and factory agents (associated with an increased risk of cancer in men) (Peired *et al.*, 2021), visceral adiposity traits (Rask-Andersen *et al.*, 2023), reproductive and hormonal risk factors (Kabat *et al.*, 2007, Lucca *et al.*, 2015), and genetic differences (Brannon *et al.*, 2012, Laskar *et al.*, 2019). A genome-wide association study (GWAS) (Laskar *et al.*, 2019) investigating sex differences of RCC reported sex-specific associations for two RCC risk SNPs (*DPF3* at 14q24 expressed in women more than men, and *EPAS1* at 2p21 expressed in men more than women), which advocates the need for more well-powered genetic studies to further elucidate these sex differences.

### 4.4.4 Rectum

As has been observed in other studies (Jackson *et al.*, 2022, White *et al.*, 2018), in UKB we observe a higher risk of cancer of the rectum (1.7-fold) in men compared to women, but not of colon cancer, even after accounting for colorectal cancer risk factors. Potential explanations for this sex difference in rectal cancer incidence may include sex hormone-related mechanisms, with some evidence of oestrogen and progesterone conferring a protective effect in women (Lin and Giovannucci, 2010, Murphy *et al.*, 2017). Although not well powered due to the rarity of the syndromes, studies reported that men with Klinefelter syndrome (an extra copy of the X chromosome) have a reduced risk of colorectal cancer

and women with Turner syndrome (only one copy of the X chromosome) have an increased risk (Lin and Giovannucci, 2010). Sex differences in visceral adiposity traits may also explain some risk differences as studies have reported interactions with obesity traits and sex in relation to colorectal cancer risk (Rask-Andersen *et al.*, 2023). Differences in screening uptake have been suggested to be another contributory factor, as men generally have lower rates of screening uptakes (both colonoscopies and sigmoidoscopies) (White *et al.*, 2018). However, if screening differences explained some excess risk for rectal cancer, we might expect to observe a somewhat similar sex differences for colon cancer. Further research is needed to explore risk factors specific to rectal cancer development in men that may explain their excess risk. Additionally, it is important to note that residual confounding from alcohol may remain, an important risk factor for colorectal cancer. The sensitivity analyses of low alcohol consumption were restricted to light drinkers as relatively few participants reported never drinking alcohol (64.0% of the cohort were light drinkers compared with 7.9% who were never drinkers).

### 4.4.5 Lung adenocarcinoma

The overall lung cancer risk was higher in men compared to women in minimally-adjusted models, but this excess risk was eliminated after adjusting for smoking and other factors. The higher risk of lung adenocarcinoma in women compared to men that we observed in the UKB has been consistently reported in other epidemiological studies, however conclusive evidence on risk factors that may explain excess risk is limited (Freedman *et al.*, 2008, Jemal *et al.*, 2018, Jemal *et al.*, 2025). Possible explanations include sex differences in chronic lung conditions such as asthma, which has a higher prevalence in women in adulthood and may be linked with increasing cancer risk with persistent damage through inflammation or trauma (Pirie *et al.*, 2016). Our sensitivity analysis in never smokers adjusted for exposure to tobacco smoke at home and outside home in participants who

identified as never smokers, however the increased risk in women remained significant. Cross sectional studies on molecular signatures in tumour tissue in The Cancer Genome Atlas have suggested marked sex differences in molecular profiles of patients with lung adenocarcinoma (Yuan *et al.*, 2016). For example, somatic mutations of the *STK11* gene were more frequent in men compared to women, whereas mutations in the *DMD* gene were more frequent in women (Yuan *et al.*, 2016).

### 4.4.6 Thyroid

We observed the risk of thyroid cancer was almost 2.7-fold higher in women compared to men; an excess risk that has been investigated in other epidemiological studies (Leclair *et al.*, 2021, Tran and Davies, 2023) previously but suggested to be not a real sex difference in incidence and instead an artefact of diagnostic differences between men and women. A cross-sectional analysis of US adults (using the National Cancer Institute's Surveillance, Epidemiology, and End Results Program data) reported the sex disparity in thyroid cancer was confined to small subclinical papillary thyroid cancers and they were identified more commonly in women than men, with its subclinical prevalence being similar (1:1) between the sexes and with men being diagnosed later in life (Tran and Davies, 2023). They suggest it could be due to women's behaviour towards accessing healthcare services more commonly than men and being referred for thyroid-ultrasonography for general concerns including fatigue and menstrual disturbances (Leclair *et al.*, 2021, Tran and Davies, 2023). There is also some evidence that hormonal and reproductive factors in women may contribute to excess risk (O'Grady *et al.*, 2024, Peterson *et al.*, 2012, Schubart *et al.*, 2021).

### 4.4.7 Strengths, Limitations and Conclusions

Notable strengths of the analyses include them being performed in one of the largest prospective cohort studies to date, being well powered for most cancer sites, with most relevant risk factors associated with respective cancers measured in the cohort, and a

literature-based approach to identifying those risk factors. The study design also allowed investigation of the most common cancer sites thus avoiding outcome selection bias and maintaining consistency in risk estimation.

Additionally, the present work represents a research contribution that complements and extends the findings of Jackson *et al*. (2022), who investigated sex differences across 21 cancer sites in the NIH-AARP Diet and Health Study of over 294,000 participants. They reported that observed sex differences persisted after adjustment for a broad set of risk factors, leading the authors to attribute the residual male predominance largely to sex-related biological mechanisms. The work presented in this chapter complements and extends these findings in three important respects. First, the present analyses had greater statistical power for several endpoints shared between the studies, including oesophageal adenocarcinoma and squamous cell carcinoma, rectal, and anal, and liver cancers, allowing for more precise relative risk estimates for these cancer sites. In addition, the current study evaluated several additional endpoints, including lung cancer histological subtypes and lymphatic and haematopoietic subtypes (and breast cancer) enabling more detailed stratified analyses. Second, the analyses addressed residual confounding, which Jackson and colleagues identified as a key limitation. Sensitivity analyses restricted to never smokers, light alcohol drinkers, and participants with normal BMI demonstrated that, for several cancers, sex differences observed in multivariable-adjusted models did not persist under these restrictions. Third, site-specific risk factors hypothesised to contribute to observed sex differences were evaluated using an *a priori*, literature-informed approach to risk factor selection, in contrast to the data-driven stepwise selection approach used by Jackson and colleagues. Taken together, these methodological extensions provide a more comprehensive investigation of sex differences in cancer risk. The findings suggest that fewer cancer sites demonstrate substantial residual sex differences after adjustment for

established risk factors than previously proposed, thereby narrowing the set of cancers for which additional biological mechanisms warrant further investigation.

Nonetheless, there are some limitations. It is likely that the analyses were subject to some residual confounding, including for example imperfectly measured smoking and alcohol consumption. This is notably relevant for cancers of the anus, gastric cardia, and oral cavity, and for leukaemia, where relative risks in men compared to women were somewhat attenuated in analyses restricted to never smokers or light drinkers (or for some cancers could not be analysed as the subgroups did not have enough cases for a robust analysis; statistical power calculations further detailed in section 3.7).

There were putative or established risk factors that were not included in the models either because they were not captured in the cohort, or because they were traits with some but limited evidence of their associations with cancer sites. These included occupations, carcinogenic infections, comorbidities, and traits (for example, metabolic syndrome, body composition measures). To address these limitations, subsequent research chapters 5 and 6 are aimed at assessing the contributions from additional risk factors to better understand the sex differences in oesophageal adenocarcinoma, lung adenocarcinoma, and cancers of the kidney and rectum.

In these pan-cancer analyses in the UK Biobank, sex differences in cancer incidence were identified that are only partially understood by known cancer risk factors. Particularly marked differences included a higher incidence of oesophageal adenocarcinoma and of cancers of the bladder, rectum, and kidney in men, and lung adenocarcinoma in women. Future research on the unexplained sex differences in cancer risk between men and women may provide insights into cancer aetiology.

# Chapter 5 Exploring Contributions of Adiposity Traits to Sex Differences in Oesophageal adenocarcinoma, Kidney, and Rectal Cancer Risk

## 5.1 Introduction

Oesophageal adenocarcinoma (OAC) and cancers of the kidney and rectum have notable sex disparities, with significantly higher incidence rates in men compared to women (Jackson *et al.*, 2022, Khan *et al.*, 2025). The excess risk is particularly pronounced for OAC, which has a four- to five-fold higher incidence in men than in women, and is approximately two-fold higher for cancers of the kidney and rectum. These disparities are observed after prospective analyses have assessed sex differences by adjusting for established risk factors hypothesized to contribute to them, such as body mass index (BMI), height, smoking, alcohol consumption, and site-specific risk factors (Jackson *et al.*, 2022, Khan *et al.*, 2025). The estimates were also consistent for OAC and kidney cancer when analyses were restricted to never smokers and people with normal BMI (Chapter 4) (Table 5.1). Therefore, it is understood that known covariates alone do not fully contribute towards the observed male predominance.

Excess adiposity is a shared risk factor for OAC and cancers of the kidney and rectum (World Cancer Research Fund, 2018). In the UK, men have consistently higher rates of overweight and obesity compared to women (Table 2.2) (NHS Digital, 2022a). Approximately 69% of adult men versus 59% of women fall into overweight or obese categories (Table 2.2) (NHS Digital, 2022a). The World Health Organization classifies people

with a BMI <25kg/m$^2$ as not overweight nor obese (or having 'normal' BMI), ≥25 to <30kg/m$^2$ as overweight and ≥30kg/m$^2$ as obese (NHS Digital, 2022a). The International Agency for Research on Cancer (IARC) report on body-fatness and cancer determined relative risks of cancers for individuals with high BMI compared to normal BMI were 4.8 (95% CI 3.0–7.7) for OAC, 1.3 (95% CI 1.3–1.4) for colorectal cancer, and 1.8 (95% CI 1.7–1.9) for kidney cancer (Lauby-Secretan *et al.*, 2016). Additionally, elevated BMI is associated with risk increases of similar magnitude in men and women (Lauby-Secretan *et al.*, 2016).

There is emerging evidence suggesting that adiposity measures beyond BMI, notably waist circumference (WC), waist-to-hip ratio (WHR), and body fat percentage, are also important predictors of risk (Keum *et al.*, 2018, Macinnis *et al.*, 2006, O'Doherty *et al.*, 2012, Rask-Andersen *et al.*, 2023, Sanikini *et al.*, 2020a, Zhang *et al.*, 2018). Although BMI is most commonly used to assess adiposity, it inadequately captures sex-specific differences in body composition and fat distribution (Bredella, 2017, Rask-Andersen *et al.*, 2023). BMI ranges considered overweight (25–29.9 kg/m$^2$) frequently fail to distinguish between lean and fat mass, particularly in men and older individuals (Romero-Corral *et al.*, 2008). Women typically have a higher body fat percentage at equivalent BMI levels, predominantly storing fat subcutaneously around the hips and thighs, whereas men are more prone to accumulating visceral abdominal fat (Bredella, 2017, Rask-Andersen *et al.*, 2023). Such distinctions in fat distribution may partly explain sex-specific differences in disease risk, supported by genome-wide association studies (GWAS) indicating sex-specific genetic differences in adiposity traits (Bredella, 2017, Kaisinger *et al.*, 2023, Pulit *et al.*, 2019, Rask-Andersen *et al.*, 2019).

Prospective studies have reported consistently positive associations between abdominal adiposity measures (WC, WHR) and OAC risk and kidney cancer risk in both men and women (Macinnis *et al.*, 2006, Murphy *et al.*, 2018, Nam *et al.*, 2019, O'Doherty *et al.*, 2012,

Parra-Soto *et al.*, 2021, Rask-Andersen *et al.*, 2023, Sanikini *et al.*, 2020a). For rectal cancer, the prospective evidence has been relatively less consistent, and weaker or null relationships have generally been observed with BMI, WC, and WHR, especially among women (Keum *et al.*, 2018, Laskar *et al.*, 2025, Murphy *et al.*, 2018, Ortega *et al.*, 2017, Parra-Soto *et al.*, 2021, Safizadeh *et al.*, 2025).

Therefore, given that several adiposity traits besides BMI are associated with risk of cancers, adiposity measurements beyond BMI differ in their distributions by sex, and that central obesity is an important risk factor hypothesized to contribute to their male predominance, this chapter is aimed at evaluating contributions of adiposity measures beyond BMI for the observed sex differences in oesophageal adenocarcinoma (OAC) and cancers of the kidney and rectum. Using UK Biobank data, this chapter covered three objectives: first, to systematically evaluate sex differences in adiposity traits in the cohort; second, to quantify associations between traits and OAC and cancers of the kidney and rectum in men and in women; and third, to quantify whether the sex differences in the risk of cancers were attenuated upon adjustment of adiposity traits. Six adiposity measures including BMI (retained as a reference) were examined: body fat percentage (BFP), waist circumference (WC), hip circumference (HC), waist-to-hip ratio (WHR), waist-to-height ratio (WHtR), and trunk fat percentage (TFP). Traits were selected *a priori* to represent complementary aspects of total and regional adiposity; although some measures were more strongly correlated with BMI than others, they were retained to allow for a comprehensive, hypothesis-agnostic assessment of whether alternative adiposity indicators added value for evaluating sex differences.

**Table 5.1 Sex differences in risk of OAC and cancers of the kidney and rectum for men compared to women (summary of results from Chapter 4; Table 4.4).**

| | **Main cohort analyses**[1] (470,720 participants) | | **Analyses restricted to participants with normal BMI (<25 kg/m²)**[2] (152,846 participants) | |
|---|---|---|---|---|
| | **Cases** (% women) | **HR (95% CI)** *p-value* | **Cases** (% women) | **HR (95% CI)** *p-value* |
| OAC | 602 (16.2) | 5.45 (4.18-7.12), p<0.0001 | 106 (24.5) | 4.71 (2.55-8.69), p<0.0001 |
| Rectum | 1841 (36.0) | 1.7 (1.47-1.96), p<0.0001 | 507 (50.3) | 1.54 (1.18-2.01), p=0.001 |
| Kidney | 1377 (35.7) | 1.77 (1.51-2.09), p<0.0001 | 280 (45.7) | 1.99 (1.40-2.83), p<0.0001 |

**Hazard ratios and 95% confidence intervals (CI) comparing risk of cancer in men to women for OAC, kidney, and rectal cancer risk.**

[1]Cox proportional hazards (PH) regression models with attained age as the underlying time scale were used to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) for the risk of cancer in men compared to women. Minimally adjusted models were run stratifying for age, fifths of Townsend deprivation index and region. Multivariable adjusted models additionally adjusted for ethnicity, qualifications, height, BMI, smoking (status and intensity), alcohol cancer site-specific risk factors.

[2]Analyses repeated in a cohort restricted to participants who had normal BMI at baseline.

# 5.2 Methods

## 5.2.1 Cohort Exclusions and Adiposity Trait Variables

Adiposity was assessed using six traits: body mass index (BMI; retained as a reference), body fat percentage (BFP), hip circumference (HC), waist circumference (WC), waist-to-hip ratio (WHR), waist-to-height ratio (WHtR), and trunk fat percentage (TFP). Cancer endpoints were oesophageal adenocarcinoma [ICD10 C15] and cancers of the rectum [C19-C20] and kidney [C64-C65] (see Chapter 3, Section 3.6, and Appendix Table 3B).

A total of 7,955 participants with missing baseline data on waist circumference, hip circumference, body fat percentage, or trunk fat percentage were excluded, resulting in a

final analytical cohort of 462,675 participants (Table 5.2). These exclusions were applied in addition to those described for Chapter 3 (see Appendix Table 3A).

BMI variable coding has been described in Chapter 3 (Section 3.9.7). BFP (%, data field 23099) and TFP (%, data field 23127) were measured in participants at recruitment as part of the bioimpedance analysis (UK Biobank, 2011). The Tanita analyser passes a low-level electrical current through the body, measuring tissue impedance to estimate relative fat and lean mass content. The conduction speed of the current is proportional to lean muscle mass due to its high-water content (approximately 70%), in contrast to fat mass, which contains minimal water (UK Biobank, 2011). While bioimpedance analysis is a practical method for estimating body composition, it has lower accuracy than MRI- or DXA-based measurements and its estimates can be influenced by hydration status, recent food or fluid intake, and skin temperature (Feng *et al.*, 2024, Ward, 2019). Central adiposity was measured using HC (cm, data field 49) and WC (cm, data field 48), alongside two derived ratios: WHR and WHtR. Waist and hip circumferences were measured manually using a calibrated, non-stretchable Seca 200 tape measure at the natural waist and the widest part of the buttocks, respectively. WHR was calculated as WC divided by HC, and WHtR as WC divided by standing height (cm, data field 50).

Results presented in this chapter are structured according to overall adiposity metrics (BMI and BFP), followed by central adiposity measures from tape measurements (HC and WC), central adiposity ratios (WHR and WHtR), and central adiposity measures from impedance analysis (TFP).

## 5.2.2 Cross-sectional Analyses of Adiposity Measures

Pairwise correlations of adiposity traits were assessed by sex to evaluate interrelatedness (Table 5.3 and 5.4). Additionally, the correlation of evaluated adiposity measures with MRI (magnetic resonance imaging) and DXA (dual-energy X-ray absorptiometry) derived

adiposity phenotypes were explored. In 2014, the UK Biobank imaging study re-invited a subsample of participants to undergo abdominal MRI and DXA (UK Biobank, 2014). In brief, participants were scanned in a Siemens MAGNETOM Aera 1.5 T MRI scanner (Siemens Healthineers, Germany) using a 6-min dual-echo Dixon Vibe protocol, providing water-and-fat-separated volumetric information for fat and muscle. Body composition analyses for MRI images were performed using AMRA Profiler Research (AMRA Medical AB, Linköping, Sweden). DXA captures whole-body composition (e.g. bone, fat and lean mass) with no extensive additional processing and analysis. Assessing correlations against imaging-based traits provided a reference for evaluating trait validity.

Category 149 was used for the MRI-based measures (UK Biobank, 2014), which were volumes (litres) of visceral adipose tissue (VAT; data field 22407), abdominal subcutaneous adipose tissue (ASAT; data field 22408) and total adipose tissue (data field 22415). These measurements are derived from abdominal MRI scans and are considered 'gold-standard' assessments of body composition. Category 124 was used for DXA-based measures (UK Biobank, 2014), which were trunk tissue fat percentage (data field 23286), android tissue fat percentage (data field 23247), gynoid tissue fat percentage (data field 23264) and whole-body tissue fat percentage (data field 23281).

### 5.2.3 Statistical Analyses

First, Cox proportional hazards models were run separately in men and women to examine associations between adiposity traits and risk of OAC and cancers of the kidney and rectum (Tables 5.6 and 5.7). Quartiles of each adiposity trait were categorised separately in men and women ('sex-stratified quartiles') to facilitate comparisons (Table 5.5). Additionally, both standard BMI categories and sex-stratified BMI quartiles were retained to allow ease of comparison with epidemiological thresholds and other traits.

Cox models were run with attained age as the underlying timescale and were cancer site-specific multivariable-adjusted (described in Chapters 3 and 4). In brief, models were stratified by age group at recruitment, region, and Townsend deprivation index, and were adjusted for ethnicity, qualification status, height, smoking, alcohol, and cancer site-specific risk factors. Site specific risk factors included gastro-oesophageal reflux disease (GORD) for OAC models; prevalent diabetes, processed meat intake, and inflammatory bowel disease (IBD) for rectal cancer; and hypertension for kidney cancer. Height was retained as an adjustment variable in all models to ensure comparability of model structure across traits. Categorical adjustments for covariates were applied to all models, except height, which was adjusted for on a continuous scale.

To explore the strength and direction of associations across increasing levels of adiposity, both p-values for trend (p-trend; testing for a linear trend across ordered quartiles) and p-values for overall association (p-heterogeneity; testing the null hypothesis of no association across adiposity quartiles compared with the reference) were calculated and reported. As the primary aim of the analyses was to assess attenuation of sex differences after adjustment for adiposity traits, no formal correction for multiple testing was applied; instead, findings were interpreted based on the magnitude of associations and overall patterns across traits.

As the initial sex-stratified analyses did not demonstrate consistent associations between adiposity traits and rectal cancer risk, subsequent analyses examining sex differences were not conducted for rectal cancer (Appendix Table 5A).

Second, Cox proportional hazards models were fitted to examine associations between sex-stratified quartiles of adiposity traits (Table 5.5) and cancer risk in a combined cohort of men and women. Using the lowest quartile in women (Q1) as the reference group, HRs

and 95% CIs were plotted against quartile medians and visually compared for cancer risk across adiposity levels in men and women on the logarithmic scale (Figures 5.1 and 5.2).

Third, cancer site-specific multivariable-adjusted Cox proportional hazards models were run with sex as the exposure variable, adjusting for each adiposity trait one at a time (Tables 5.8 and 5.9). Attenuation of HRs (and 95% CIs) was assessed by comparing results to a model with no adiposity-trait and a BMI-adjusted model. Adiposity traits were adjusted for both in a continuous and categorical format; categorical adjustment used trait quartiles that were not sex-stratified to maintain comparability across sexes. The percentage change in the natural logarithm of the hazard ratio (%Δ ln(HR)) across models and Z-scores were examined to quantify the impact of adiposity trait adjustment on effect estimates (described in Section 3.8, Appendix Tables 5B and 5C).

References to the results figures and tables have been made throughout the methods section for ease of reference.

# 5.3 Results

## 5.3.1 UK Biobank Participants

The final study cohort comprised 462,765 participants aged 37 to 73 years at recruitment, with 249,602 women (53.9%) and 213,163 men (46.1%). The average age at recruitment was 56 years (SD 8) for both men and women. The cohort was followed up on average for 10.5 (SD 2.2) years; 10.6 (SD 2.1) years for women and 10.4 (SD 2.4) years for men. 656 cases (104 in women; 15.9%) of oesophageal adenocarcinoma, 1,806 cases (36.2% in women) of rectal cancer, and 1,349 cases (35.7% in women) of kidney cancer were identified. The mean body mass index (BMI) for the entire cohort was 27.4 (SD 4.8); 27.1 (SD 5.2) for women and 27.9 (SD 4.2) for men. At recruitment, 32.4% of the cohort population had a BMI <25 (of which 65.3% were women), 42.9% had a BMI ranging between 25-29.9 (of which 46.6% were

women), 17.7% had a BMI ranging from 30-34.9 (of which 48.1% were women), and 7% had a BMI ≥35 (of which 61.6% were women). Across BMI categories, higher levels of alcohol consumption were markedly more common among men than women, with the greatest sex differences observed in the highest BMI groups (Table 5.2).

### 5.3.2 Cross-sectional Associations Between Adiposity Traits

In men, strong positive correlations were observed between most anthropometric adiposity traits (Table 5.3). Body mass index (BMI) was most strongly correlated with waist circumference (WC; r = 0.88), waist-to-height ratio (WHtR; 0.88), hip circumference (HC; 0.81) and body fat percentage (BFP; 0.80), and least with waist-to-hip ratio (WHR; 0.59). BFP and trunk fat percentage (TFP) were also strongly correlated (0.99). WC was strongly correlated with all traits (from 0.94 for WHtR to 0.76 for TFP). HC was most strongly correlated with WC (0.82) and least with WHR (0.31). All correlations were statistically significant at $p<0.001$.

Additionally, in men, for MRI-derived adiposity measures, visceral adipose tissue volume (VAT) was moderately correlated with HC (0.49), BMI (0.38) and WC (0.33) and weakly correlated with other traits (WHtR, 0.14; BFP; 0.10, TFP; 0.08; and WHR, 0.06). Abdominal subcutaneous adipose tissue volume (ASAT) was strongly correlated with all traits (0.76 for BMI to 0.63 for TFP) and moderately correlated with WHR (0.49). Total adipose tissue volume had similar associations, strong correlations with traits (0.77 for BMI to 0.67 for TFP) except WHR for which the correlation was moderate (0.55). VAT was moderately correlated with ASAT (0.28) and total adipose tissue volume (0.26), whereas ASAT and total adiposity tissue volume were highly correlated (0.96). DXA-derived adiposity measures were highly correlated within regional and whole-body fat percentage traits. Anthropometric adiposity traits (BMI, WC, and WHtR) had moderate-to-strong correlations with DXA-derived fat

percentage measures and moderate correlations with WHR. All reported correlations were statistically significant ($p < 0.001$).

In women, strong positive correlations were observed between most anthropometric adiposity traits (Table 5.4). BMI was most strongly associated with HC (0.89), WC (0.88), WHtR (0.87) and BFP (0.85), and least with WHR (0.46). BFP and TFP were also strongly correlated (0.98). WC was strongly correlated with all traits (from 0.97 for WHtR to 0.75 for WHR and TFP). HC was most strongly correlated with BMI (0.88) and WC (0.83) and least with WHR (0.24). All correlations were statistically significant at $p<0.001$.

For MRI-derived adiposity measures, in women, VAT was moderately correlated with HC (0.42), WC (0.39), and BMI (0.37), and weakly correlated with other traits (BFP, 0.24; TFP, 0.24; WHtR, 0.23; and WHR 0.18). ASAT was strongly correlated with adiposity traits (0.81 for BMI to 0.70 for TFP) and moderately correlated with WHR (0.39). Total adipose tissue volume had similar associations, strong correlations with traits (0.82 for BMI to 0.73 for TFP) except WHR for which the correlation was moderate (0.37). VAT was moderately correlated with ASAT (0.35) and total adipose tissue volume (0.36), whereas ASAT and total adiposity tissue volume were highly correlated (0.96). DXA-derived adiposity measures were highly correlated within regional and whole-body fat percentage traits. Anthropometric adiposity traits (BMI, WC, and WHtR) had moderate-to-strong correlations with DXA-derived fat percentage measures, whereas waist-to-hip ratio has consistently weaker associations. All reported correlations were statistically significant at $p<0.001$.

## 5.3.3 Sex-stratified Associations for Adiposity Traits and Cancers

### Oesophageal adenocarcinoma [C15]

All adiposity traits examined (Table 5.5) were positively associated with risk of oesophageal adenocarcinoma (OAC) in both men and women (Table 5.6). In men, the largest

associations observed were for BFP (Q4 vs Q1 HR 2.36; 95% CI 1.77-3.16), followed by WC (HR 2.31; 95% CI 1.75-3.05), TFP (HR 2.21; 95% CI 1.66-2.95) WHR (HR 2.19; 95% CI 1.65-2.91), WHtR (HR 2.19; 95% CI 1.64-2.93) and BMI (HR 2.06; 95% CI 1.58-2.70), and HC (HR 1.59; 95% CI 1.25-2.03). In women, the largest associations observed were for WC (Q4 vs Q1 HR 3.60; 95% CI 1.72-7.52), followed by WHtR (HR 2.78, 95% CI 1.35-5.73), WHR (HR 2.72, 95% CI 1.33-5.53), BMI (HR 2.37; 95% CI 1.27-4.46), BFP (HR 2.01; 95% CI 1.09-3.69) and TFP (HR 1.91; 95% CI 1.07-3.40), and HC (HR 1.84; 95% CI 1.03-3.29).

## Rectum [C19-C20]

Associations between adiposity traits and rectal cancer risk varied by sex (Appendix Table 5A). In men, associations were observed for WHR (Q4 vs Q1 HR 1.32; 95% CI 1.10-1.58), WC (HR 1.31; 95% CI 1.10-1.57) and WHtR (HR 1.23; 95% CI 1.03-1.47), and no other adiposity trait. In women, only HC was associated with lower risk (HR 0.75; 95% CI 0.60-0.94), and no other adiposity trait. Based on these results, no further analyses were performed to assess contributions from adiposity traits in relation to sex differences in rectal cancer risk.

## Kidney [C64-C65]

All adiposity traits examined (Table 5.5) were positively associated with risk of kidney cancer in both men and women (Table 5.7). In men, the largest associations observed were for WHtR (Q4 vs Q1 HR 1.94; 95% CI 1.57-2.40), followed by BMI (HR 1.87; 95% CI 1.53-2.28), WHR (HR 1.81; 95% CI 1.47-2.22), WC (HR 1.80; 95% CI 1.46-2.20), TFP (HR 1.59; 95% CI 1.30-1.95) and BFP (HR 1.55; 95% CI 1.27-1.90), and HC (HR 1.51; 95% CI 1.25-1.84). In women, the largest associations observed were for WHtR (Q4 vs Q1 HR 2.31; 95% CI 1.71–3.13), followed by WHR (HR 2.25; 95% CI 1.67–3.03), BMI (HR 2.17; 95% CI 1.63–2.90), WC (HR 2.16; 95% CI 1.62–2.87), BFP (HR 2.01; 95% CI 1.50-2.69) and HC (HR 1.71; 95% CI 1.32-2.21), and TFP (HR 1.68; 95% CI 1.27-2.22).

## 5.3.4 Sex Differences in Cancer Risk by Adiposity Traits

### Oesophageal adenocarcinoma [C15]

Higher levels of overall and central adiposity were associated with increased risk of OAC in both men and women, with associations consistently stronger in men compared to women (Figure 5.1). Visual comparisons of sex-stratified quartile-based hazard ratios showed that associations for WHR and WC were stronger than those observed for BMI, especially in men.

Consistent with these patterns, adjustment for adiposity traits, specifically WC and WHR, partially attenuated the sex difference in OAC risk (Table 5.8). In multivariable-adjusted Cox models comparing men with women, the hazard ratio for OAC in a model without adiposity trait adjustment was 5.55 (95% CI 4.24–7.25). Adjusting for WHR yielded the largest attenuation (HR 3.54; 95% CI 2.59-4.83), followed by WC (HR 4.40; 95% CI 3.33-5.81), WHtR (HR 4.48; 95% CI 3.40-5.89) and BMI (HR 5.12; 95% CI 3.91-6.71). The results from models with continuous adjustment were consistent with those from models with categorical adjustments for traits. Corresponding Z-scores decreased following adjustment for WHR (Z=7.9) and WC (Z=10.4), compared with the model without adiposity trait adjustment (Z=12.6). The percentage change in the natural logarithm of the hazard ratio relative to the unadjusted model was largest for WHR (26.6%), followed by WC (13.5%), WHtR (12.5%), and BMI (4.7%) (Appendix Table 5B).

Adjustment for other adiposity traits was associated with higher hazard ratio estimates compared with the model without adiposity trait adjustment. Specifically, adjustment for BFP had a HR 8.35 (95% CI 6.08-11.47), followed by TFP (HR 6.65; 95% CI 5.04-8.78) and HC (HR 6.11; 95% CI 4.65-8.02). This is consistent with the higher distribution of these adiposity traits in women (Table 5.5).

## Kidney [C64-C65]

Higher levels of overall and central adiposity were associated with increased risk of kidney cancer in both men and women, with associations consistently stronger in men compared to women (Figure 5.2). Visual comparisons of sex-stratified quartile-based hazard ratios showed that associations for WHR and WC were stronger than those observed for BMI, especially in men.

Consistent with these patterns, adjustment for adiposity traits, specifically WC and WHR, strongly attenuated the sex difference in kidney cancer risk (Table 5.9). In multivariable-adjusted Cox models comparing men with women, the hazard ratio for kidney cancer in a model without adiposity trait adjustment was 1.82 (95% CI 1.54-2.15). Adjusting for WHR yielded the greatest attenuation (HR 1.36; 95% CI 1.13-1.65), followed by WC (HR 1.58; 95% CI 1.33-1.88), WHtR (HR 1.59; 95% CI 1.34-1.88) and BMI (1.74; 95% CI 1.47-2.05). The results from models with continuous adjustment were consistent with those from models with categorical adjustments for traits. Corresponding Z-scores decreased following adjustment for WHR (Z=3.2) and WC (Z=5.3), compared with the model without adiposity trait adjustment (Z=7.2). The percentage change in the natural logarithm of the hazard ratio relative to the unadjusted model was largest for WHR (48.5%), followed by WC (23.4%), WHtR (22.4%), and BMI (7.4%) (Appendix Table 5C).

Adjustment for other adiposity traits was associated with higher hazard ratio estimates compared with the model without adiposity trait adjustment. Specifically, adjustment for BFP had a HR 2.43 (95% CI 1.99-2.97), followed by TFP (HR 2.11; 95% CI 1.77-2.52) and HC (HR 1.95; 95% CI 1.65-2.31). This is consistent with the higher distribution of these adiposity traits in women (Table 5.5).

**Table 5.2 Distribution of 462,765 UK Biobank participants (53.9% women) by body mass index (BMI) measures at baseline and selected sociodemographic factors.**

| | BMI <25 | BMI 25–29.9 | BMI 30–34.9 | BMI ≥35 |
|---|---|---|---|---|
| **Total cohort** | **149,787 (65.3)** | **198,540 (46.6)** | **82,097 (48.1)** | **32,341 (61.6)** |
| **Age at recruitment (years)** | | | | |
| <45 | 19,168 (66.4) | 19,513 (42.5) | 7,536 (45.0) | 3,234 (63.0) |
| 45-49 | 22,987 (68.1) | 25,426 (45.7) | 10,080 (48.0) | 4,423 (62.6) |
| 50-54 | 24,031 (67.9) | 29,277 (48.2) | 12,844 (49.3) | 5,545 (62.7) |
| 55-59 | 26,977 (66.5) | 35,511 (48.2) | 15,184 (48.7) | 6,381 (62.3) |
| 60-64 | 32,664 (64.0) | 49,281 (47.8) | 20,434 (49.0) | 7,664 (60.9) |
| >65 | 23,960 (59.4) | 39,532 (45.0) | 16,019 (46.8) | 5,094 (59.0) |
| **Townsend Deprivation Index (quintiles of most affluent to most deprived)** | | | | |
| Most affluent | 32,351 (66.8) | 41,546 (45.4) | 14,734 (45.9) | 4,432 (59.8) |
| 2nd | 31,068 (66.6) | 41,069 (46.2) | 15,432 (47.3) | 5,147 (60.5) |
| 3rd | 30,069 (66.9) | 40,505 (47.4) | 16,195 (48.0) | 5,844 (60.6) |
| 4th | 29,310 (65.4) | 39,060 (47.6) | 16,914 (49.2) | 7,127 (61.9) |
| Most deprived | 26,817 (60.0) | 36,121 (46.3) | 18,714 (49.4) | 9,741 (63.5) |
| Unknown* | 172 (64.5) | 239 (41.8) | 108 (48.1) | 50 (54.0) |
| **Region** | | | | |
| London | 23,550 (65.2) | 25,794 (46.6%) | 9,997 (50.4%) | 4,111 (64.3) |
| North-West England | 21,553 (64.3) | 31,113 (45.9) | 13,493 (47.2) | 5,606 (59.4) |
| North-Eastern England | 15,450 (66.4) | 23,865 (47.3) | 10,352 (48.0) | 4,060 (60.2) |
| Yorkshire & the Humber | 21,551 (65.3) | 30,004 (46.8) | 12,419 (47.8) | 4,788 (61.8) |
| West Midlands | 11,970 (62.6) | 18,083 (44.0) | 7,950 (45.1) | 3,126 (59.9) |
| East Midlands | 10,202 (64.8) | 13,547 (46.9) | 5,583 (48.5) | 2,119 (62.1) |
| South-East England | 15,070 (66.2) | 16,924 (46.3) | 6,133 (49.1) | 2,156 (64.3) |
| South-West England | 14,304 (66.1) | 16,830 (47.3) | 6,459 (49.3) | 2,415 (63.0) |
| Wales | 5,321 (66.2) | 8,215 (47.1) | 3,889 (47.5) | 1,681 (61.0) |
| Scotland | 10,816 (66.4) | 14,165 (48.2) | 5,822 (48.3) | 2,279 (62.5) |
| **Education status (highest level qualifications)** | | | | |
| Prof Q/NVQ/HND/HNC/Degree or other professional qualification | 95,843 (63.1) | 117,518 (42.9) | 44,892 (45.2) | 16,880 (60.2) |

| | | | | |
|---|---|---|---|---|
| A level | 8,841 (69.4) | 10,607 (49.6) | 4,125 (49.5) | 1,638 (64.5) |
| O level/GCSE/CSE | 24,246 (74.2) | 32,837 (56.1) | 13,970 (56.0) | 5,790 (67.0) |
| Unknown* | 20,857 (63.3) | 37,578 (48.9) | 19,110 (48.6) | 8,033 (60.2) |
| **Ethnicity** | | | | |
| White | 141,688 (65.7) | 186,954 (46.6) | 76,786 (47.6) | 30,025 (60.8) |
| Mixed Race | 976 (71.1) | 1,075 (55.0) | 486 (57.0) | 216 (68.5) |
| Asian or Asian British | 3,869 (53.6) | 4,703 (42.8) | 1,607 (49.7) | 496 (66.5) |
| Black or Black British | 1,373 (54.6) | 3,094 (48.7) | 1,921 (62.4) | 1,072 (79.3) |
| Other | 1,252 (66.0) | 1,798 (47.8) | 839 (53.3) | 346 (74.6) |
| Unknown* | 629 (52.8) | 916 (38.1) | 458 (47.6) | 186 (54.3) |
| **Alcohol consumption, grams per day** | | | | |
| <1 g/day | 15,192 (76.8) | 19,519 (68.6) | 10,578 (70.8) | 6,354 (79.9) |
| >=1-<10 g/day | 52,137 (75.7) | 58,777 (62.0) | 23,945 (63.0) | 9,990 (70.5) |
| >=10-<20 g/day | 36,800 (66.5) | 43,770 (47.1) | 14,877 (46.4) | 4,658 (56.7) |
| >=20-<30 g/day | 17,845 (52.3) | 27,218 (30.4) | 9,843 (29.5) | 2,798 (39.7) |
| 30+ g/day | 16,005 (29.5) | 34,005 (14.3) | 14,875 (14.2) | 4,107 (18.8) |
| Non-drinkers | 10,883 (68.7) | 14,081 (57.9) | 7,404 (62.6) | 4,145 (74.5) |
| Unknown* | 925 (70.7) | 1,170 (57.6) | 575 (59.0) | 289 (67.1) |
| **Smoking status and intensity, cigarettes per day** | | | | |
| Current, <15 cigs/day | 5,753 (64.6) | 5,639 (51.4) | 1,937 (54.3) | 695 (68.6) |
| Current, >=15 cigs/day | 6,746 (49.0) | 7,730 (40.8) | 3,400 (43.8) | 1,365 (56.4) |
| Current, unknown cigs/day | 4,519 (48.4) | 7,065 (28.0) | 2,783 (28.3) | 956 (38.9) |
| Former, <15 cigs/day | 9,478 (69.2) | 12,442 (51.8) | 4,390 (54.7) | 1,523 (70.3) |
| Former, >=15 cigs/day | 15,149 (59.5) | 32,786 (37.1) | 17,545 (35.4) | 7,158 (48.4) |
| Former, unknown cigs/day | 18,810 (66.7) | 25,449 (43.6) | 9,995 (42.7) | 3,622 (58.2) |
| Never | 88,675 (67.7) | 106,325 (51.0) | 41,504 (55.4) | 16,764 (68.7) |
| Unknown* | 657 (57.2) | 1,104 (45.9) | 543 (46.0) | 258 (58.5) |

*Counts represent the total number of participants in each subgroup, with the (%) indicating the proportion in each category who were women.*
**This category included participants who responded either 'Do not know' or 'prefer not to say'.*

**Table 5.3 Pearson correlation coefficients between baseline adiposity traits in 213,163 men from the UK Biobank.**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| **(1) BMI** | 1.00 | | | | | | |
| **(2) Body fat percentage** | 0.80 | 1.00 | | | | | |
| **(3) Hip circumference** | 0.81 | 0.66 | 1.00 | | | | |
| **(4) Waist circumference** | 0.88 | 0.79 | 0.82 | 1.00 | | | |
| **(5) Waist-to-hip ratio** | 0.59 | 0.62 | 0.31 | 0.79 | 1.00 | | |
| **(6) Waist-to-height ratio** | 0.88 | 0.82 | 0.72 | 0.94 | 0.81 | 1.00 | |
| **(7) Trunk fat percentage** | 0.76 | 0.99 | 0.64 | 0.76 | 0.60 | 0.79 | 1.00 |

All p-values were <0.001.

**Table 5.4 Pearson correlation coefficients between baseline adiposity traits in 249,602 women from the UK Biobank.**

| | **(1)** | **(2)** | **(3)** | **(4)** | **(5)** | **(6)** | **(7)** |
|---|---|---|---|---|---|---|---|
| **(1) BMI** | 1.00 | | | | | | |
| **(2) Body fat percentage** | 0.85 | 1.00 | | | | | |
| **(3) Hip circumference** | 0.89 | 0.81 | 1.00 | | | | |
| **(4) Waist circumference** | 0.88 | 0.82 | 0.83 | 1.00 | | | |
| **(5) Waist-to-hip ratio** | 0.46 | 0.46 | 0.24 | 0.75 | 1.00 | | |
| **(6) Waist-to-height ratio** | 0.87 | 0.80 | 0.77 | 0.97 | 0.75 | 1.00 | |
| **(7) Trunk fat percentage** | 0.76 | 0.98 | 0.76 | 0.75 | 0.41 | 0.70 | 1.00 |

All p-values were <0.001.

**Table 5.5 Baseline adiposity traits categorised into sex-stratified quartiles in 462,765 UK Biobank participants.**

| | Women | | | Men | | |
|---|---|---|---|---|---|---|
| **Quartile** | **N** | **Mean [SD]** | **Median** | **N** | **Mean [SD]** | **Median** |
| **Body mass index (BMI)** | | | | | | |
| 1 | 64739 | 21.69 [1.39] | 22.0 | 53976 | 23.23 [1.44] | 23.6 |
| 2 | 60925 | 24.84 [0.75] | 24.8 | 53322 | 26.22 [0.66] | 26.2 |
| 3 | 61916 | 27.79 [1.03] | 27.7 | 53749 | 28.65 [0.80] | 28.6 |
| 4 | 62022 | 34.21 [4.19] | 33.0 | 52116 | 33.53 [3.40] | 32.5 |
| **Body fat percentage (BFP)** | | | | | | |
| 1 | 63056 | 27.66 [3.94] | 28.4 | 54165 | 18.44 [4.15] | 19.5 |
| 2 | 62532 | 32.35 [1.78] | 32.3 | 53291 | 23.23 [1.46] | 23.2 |
| 3 | 61617 | 36.29 [1.71] | 36.3 | 53500 | 26.81 [1.49] | 26.8 |
| 4 | 61939 | 43.19 [4.49] | 42.0 | 51815 | 33.05 [4.41] | 31.9 |
| **Hip circumference (HC; cm)** | | | | | | |
| 1 | 64649 | 92.28 [3.36] | 93.0 | 64057 | 95.61 [3.12] | 96.0 |
| 2 | 68764 | 99.51 [1.70] | 100.0 | 51747 | 101.50 [1.12] | 102.0 |
| 3 | 53793 | 105.25 [1.68] | 105.0 | 44864 | 105.38 [1.12] | 105.0 |
| 4 | 62396 | 117.28 [8.41] | 115 | 52495 | 113.29 [6.19] | 111.0 |
| **Waist circumference (WC; cm)** | | | | | | |
| 1 | 63057 | 70.53 [3.62] | 71.0 | 54636 | 84.03 [4.36] | 85.0 |
| 2 | 66567 | 79.46 [2.29] | 79.0 | 56639 | 93.07 [1.98] | 93.0 |
| 3 | 60116 | 87.68 [2.56] | 88.0 | 49150 | 99.76 [1.98] | 100.0 |
| 4 | 59862 | 102.12 [8.45] | 100.0 | 52738 | 111.87 [7.98] | 109.1 |
| **Waist-to-hip ratio (WHR)** | | | | | | |
| 1 | 63005 | 0.78 [0.04] | 0.79 | 53942 | 0.87 [0.05] | 0.88 |
| 2 | 62509 | 0.82 [0.02] | 0.82 | 53256 | 0.91 [0.02] | 0.91 |
| 3 | 61959 | 0.86 [0.03] | 0.86 | 53143 | 0.95 [0.03] | 0.95 |
| 4 | 62397 | 0.91 [0.04] | 0.90 | 53289 | 1.02 [0.04] | 1.01 |
| **Waist-to-height ratio (WHtR)** | | | | | | |
| 1 | 62772 | 0.43 [0.02] | 0.44 | 53299 | 0.48 [0.02] | 0.48 |
| 2 | 62031 | 0.49 [0.01] | 0.49 | 53478 | 0.53 [0.01] | 0.53 |
| 3 | 62443 | 0.54 [0.02] | 0.54 | 53333 | 0.57 [0.01] | 0.57 |
| 4 | 62356 | 0.63 [0.05] | 0.62 | 53053 | 0.64 [0.05] | 0.63 |
| **Trunk fat percentage (TFP)** | | | | | | |
| 1 | 63243 | 23.97 [4.35] | 25.1 | 53645 | 19.12 [3.86] | 20.2 |
| 2 | 62683 | 31.98 [1.55] | 32.0 | 53321 | 25.98 [1.26] | 26.0 |
| 3 | 61548 | 36.99 [1.43] | 37.0 | 53367 | 30.04 [1.17] | 30.0 |
| 4 | 62128 | 43.68 [3.42] | 42.9 | 52830 | 35.67 [2.99] | 34.9 |

**Table 5.6 Hazard ratios (HRs) and 95% confidence intervals (CIs) for oesophageal adenocarcinoma [ICD-10 C15], in 249,602 women and 213,163 men according to adiposity traits in the UK Biobank.**

| **Adiposity traits** | **Categories** ***Q1 (reference)*** | **Women** **HR (95% CI)** ***104 cases*** | **Men** **HR (95% CI)** ***552 cases*** |
|---|---|---|---|
| **BMI*** (WHO categories; kg/m$^2$) | 25-29.9 | 1.44 (0.88–2.36) | 1.64 (1.26–2.12) |
| | 30-34.9 | 1.65 (0.92–2.98) | 2.06 (1.55–2.74) |
| | ≥35 | 3.20 (1.71–5.96) | 2.40 (1.66–3.45) |
| | *p-trend* | <0.001 | <0.001 |
| | *p-heterogeneity* | 0.003 | <0.001 |
| **BMI** (quartiles; kg/m$^2$) | Q2 | 1.37 (0.70–2.71) | 1.30 (0.98–1.74) |
| | Q3 | 1.93 (1.03–3.64) | 1.85 (1.41–2.42) |
| | Q4 | 2.37 (1.27–4.46) | 2.06 (1.58–2.70) |
| | *p-trend* | 0.004 | <0.001 |
| | *p-heterogeneity* | 0.032 | <0.001 |
| **Body fat percentage** (BFP) | Q2 | 1.12 (0.57–2.17) | 1.40 (1.02–1.93) |
| | Q3 | 1.41 (0.75–2.64) | 2.16 (1.61–2.90) |
| | Q4 | 2.01 (1.09–3.69) | 2.36 (1.77–3.16) |
| | *p-trend* | 0.018 | <0.001 |
| | *p-heterogeneity* | 0.057 | <0.001 |
| **Hip circumference** (HC; cm) | Q2 | 1.12 (0.60–2.08) | 1.15 (0.89–1.49) |
| | Q3 | 1.69 (0.93–3.08) | 1.54 (1.20–1.98) |
| | Q4 | 1.84 (1.03–3.29) | 1.59 (1.25–2.03) |
| | *p-trend* | 0.016 | <0.001 |
| | *p-heterogeneity* | 0.091 | <0.001 |
| **Waist circumference** (WC; cm) | Q2 | 2.12 (0.98–4.57) | 1.52 (1.14–2.04) |
| | Q3 | 2.70 (1.28–5.72) | 1.73 (1.29–2.32) |
| | Q4 | 3.60 (1.72–7.52) | 2.31 (1.75–3.05) |
| | *p-trend* | <0.001 | <0.001 |
| | *p-heterogeneity* | 0.005 | <0.001 |

| | | | |
|---|---|---|---|
| **Waist-to-hip ratio** (WHR) | Q2 | 2.28 (1.10–4.73) | 1.42 (1.05–1.93) |
| | Q3 | 2.25 (1.09–4.64) | 1.54 (1.14–2.07) |
| | Q4 | 2.72 (1.33–5.53) | 2.19 (1.65–2.91) |
| | *p-trend* | 0.008 | <0.001 |
| | *p-heterogeneity* | 0.051 | <0.001 |
| **Waist-to-height ratio** (WHtR) | Q2 | 1.61 (0.75–3.48) | 1.27 (0.93–1.74) |
| | Q3 | 2.89 (1.43–5.87) | 1.95 (1.46–2.60) |
| | Q4 | 2.78 (1.35–5.73) | 2.19 (1.64–2.93) |
| | *p-trend* | 0.001 | <0.001 |
| | *p-heterogeneity* | 0.008 | <0.001 |
| **Trunk fat percentage** (TFP) | Q2 | 1.35 (0.74–2.48) | 1.51 (1.11–2.06) |
| | Q3 | 0.80 (0.41–1.57) | 2.08 (1.56–2.78) |
| | Q4 | 1.91 (1.07–3.40) | 2.21 (1.66–2.95) |
| | *p-trend* | 0.130 | <0.001 |
| | *p-heterogeneity* | 0.010 | <0.001 |

All Cox models were stratified by age at recruitment, deprivation index, and region, and adjusted for education, ethnicity, height, smoking status, alcohol intake, and prevalent gastro-oesophageal reflux disease.

*The BMI (WHO) reference category was <25 kg/m$^2$.

**Table 5.7 Hazard ratios (HRs) and 95% confidence intervals (CIs) for kidney cancer [ICD-10 C64-C65], in 249,602 women and 213,163 men according to adiposity traits in the UK Biobank.**

| Adiposity Trait | Categories<br>*Q1 (reference)* | Women<br>HR (95% CI)<br>*481 cases* | Men<br>HR (95% CI)<br>*868 cases* |
|---|---|---|---|
| **BMI**<br>(WHO categories; kg/m$^2$) | 25-29.9 | 1.31 (1.04–1.64) | 1.23 (1.02–1.49) |
| | 30-34.9 | 1.46 (1.11–1.91) | 1.69 (1.36–2.09) |
| | ≥35 | 2.15 (1.59–2.92) | 1.98 (1.50–2.63) |
| | *p-trend* | <0.001 | <0.001 |
| | *p-heterogeneity* | <0.001 | <0.001 |
| **BMI**<br>(quartiles; kg/m$^2$) | Q2 | 1.45 (1.08–1.97) | 1.17 (0.95–1.45) |
| | Q3 | 1.69 (1.26–2.26) | 1.40 (1.14–1.72) |
| | Q4 | 2.17 (1.63–2.90) | 1.87 (1.53–2.28) |
| | *p-trend* | <0.001 | <0.001 |
| | *p-heterogeneity* | <0.001 | <0.001 |
| **Body fat percentage**<br>(BFP) | Q2 | 1.39 (1.02–1.88) | 1.15 (0.93–1.42) |
| | Q3 | 1.63 (1.21–2.19) | 1.16 (0.94–1.44) |
| | Q4 | 2.01 (1.50–2.69) | 1.55 (1.27–1.90) |
| | *p-trend* | <0.001 | <0.001 |
| | *p-heterogeneity* | <0.001 | <0.001 |
| **Hip circumference**<br>(HC; cm) | Q2 | 1.13 (0.86–1.48) | 1.21 (0.99–1.47) |
| | Q3 | 1.16 (0.87–1.55) | 1.37 (1.12–1.67) |
| | Q4 | 1.71 (1.32–2.21) | 1.51 (1.25–1.84) |
| | *p-trend* | <0.001 | <0.001 |
| | *p-heterogeneity* | <0.001 | <0.001 |
| **Waist circumference**<br>(WC; cm) | Q2 | 1.39 (1.03–1.88) | 1.03 (0.82–1.28) |
| | Q3 | 1.58 (1.18–2.13) | 1.43 (1.15–1.76) |
| | Q4 | 2.16 (1.62–2.87) | 1.80 (1.46–2.20) |
| | *p-trend* | <0.001 | <0.001 |
| | *p-heterogeneity* | <0.001 | <0.001 |

| | | | |
|---|---|---|---|
| **Waist to hip ratio** (WHR) | Q2 | 1.60 (1.17–2.18) | 1.08 (0.86–1.35) |
| | Q3 | 1.87 (1.38–2.54) | 1.42 (1.15–1.76) |
| | Q4 | 2.25 (1.67–3.03) | 1.81 (1.47–2.22) |
| | *p-trend* | <0.001 | <0.001 |
| | *p-heterogeneity* | <0.001 | <0.001 |
| **Waist to height ratio** (WHtR) | Q2 | 1.56 (1.14–2.13) | 1.29 (1.03–1.61) |
| | Q3 | 1.68 (1.24–2.29) | 1.37 (1.10–1.71) |
| | Q4 | 2.31 (1.71–3.13) | 1.94 (1.57–2.40) |
| | *p-trend* | <0.001 | <0.001 |
| | *p-heterogeneity* | <0.001 | <0.001 |
| **Trunk fat percentage** (TFP) | Q2 | 1.11 (0.82–1.49) | 1.23 (0.99–1.52) |
| | Q3 | 1.49 (1.13–1.97) | 1.23 (1.00–1.52) |
| | Q4 | 1.68 (1.27–2.22) | 1.59 (1.30–1.95) |
| | *p-trend* | <0.001 | <0.001 |
| | *p-heterogeneity* | <0.001 | <0.001 |

All Cox models were stratified by age at recruitment, deprivation index, region, and adjusted for qualifications, ethnicity, height, smoking status, alcohol, and prevalent hypertension.

*The BMI (WHO) reference category was <25 kg/m$^2$.

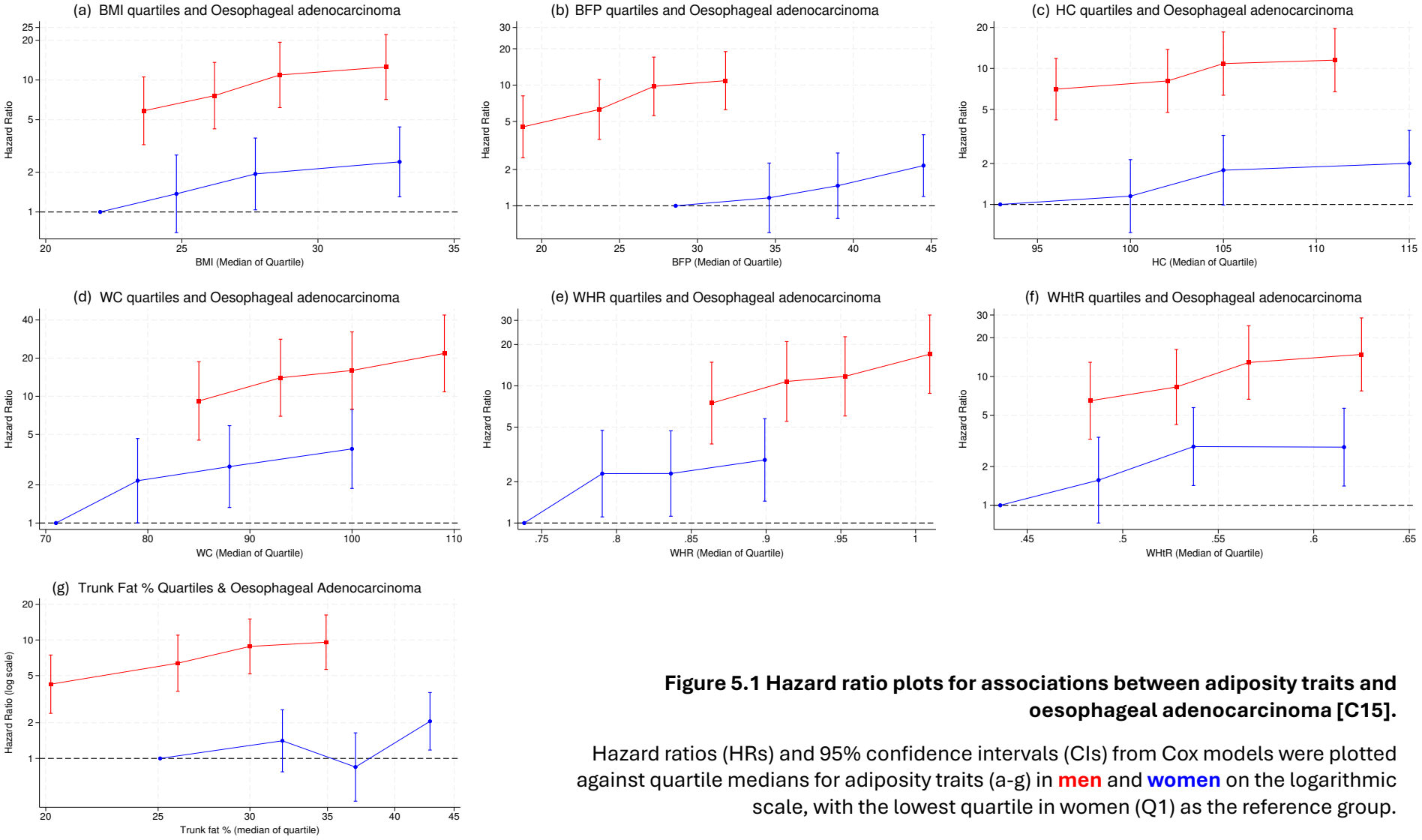



**Figure 5.1 Hazard ratio plots for associations between adiposity traits and oesophageal adenocarcinoma [C15].**

Hazard ratios (HRs) and 95% confidence intervals (CIs) from Cox models were plotted against quartile medians for adiposity traits (a-g) in **men** and **women** on the logarithmic scale, with the lowest quartile in women (Q1) as the reference group.

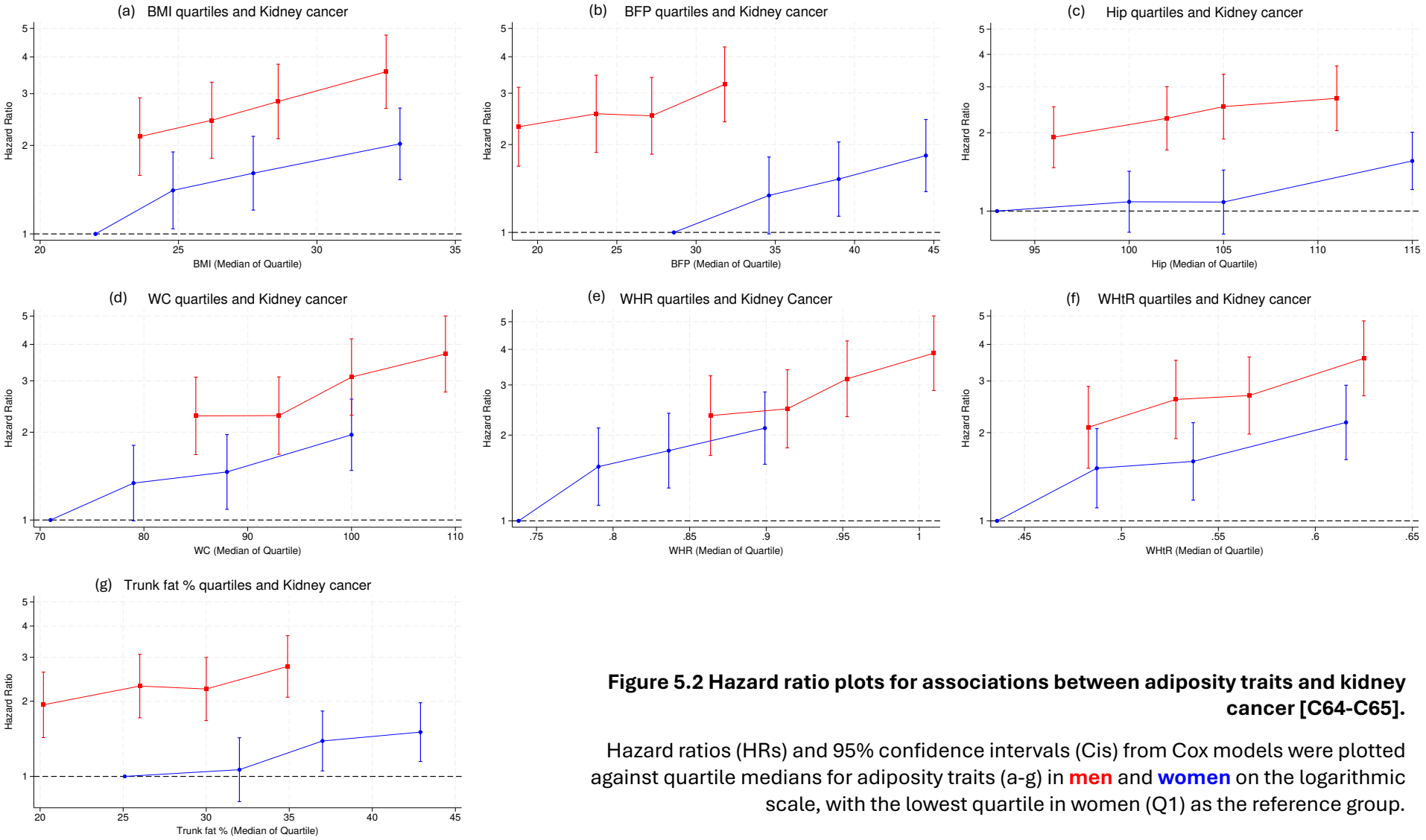



**Figure 5.2 Hazard ratio plots for associations between adiposity traits and kidney cancer [C64-C65].**

Hazard ratios (HRs) and 95% confidence intervals (Cis) from Cox models were plotted against quartile medians for adiposity traits (a-g) in **men** and **women** on the logarithmic scale, with the lowest quartile in women (Q1) as the reference group.

**Table 5.8 Hazard ratios (95% confidence intervals) comparing men with women for oesophageal adenocarcinoma [ICD-10 C15], adjusted for one adiposity trait at a time.**

| Adiposity Trait | Continuous adjustment | | | Categorical adjustment | | |
|---|---|---|---|---|---|---|
| | HR (95% CI) | p-value | Z-score | HR (95% CI) | p-value | Z-score |
| *No adiposity trait adjustment* | *5.55 (4.24-7.25)* | *<0.001* | *12.56* | *5.55 (4.24-7.25)* | *<0.001* | *12.56* |
| BMI* (kg/m$^2$) | 5.26 (4.02–6.89) | <0.001 | 12.10 | 5.12 (3.91–6.71) | <0.001 | 11.87 |
| Body fat percentage (BFP) | 10.03 (7.36–13.66) | <0.001 | 14.61 | 8.35 (6.08–11.47) | <0.001 | 13.11 |
| Hip circumference (HC; cm) | 6.34 (4.83–8.33) | <0.001 | 13.29 | 6.11 (4.65–8.02) | <0.001 | 13.04 |
| Waist circumference (WC; cm) | 4.48 (3.41–5.89) | <0.001 | 10.79 | 4.40 (3.33–5.81) | <0.001 | 10.44 |
| Waist-to-hip ratio (WHR) | 3.33 (2.48–4.48) | <0.001 | 7.97 | 3.54 (2.59–4.83) | <0.001 | 7.94 |
| Waist-to-height ratio (WHtR) | 4.47 (3.41–5.88) | <0.001 | 10.77 | 4.48 (3.40–5.89) | <0.001 | 10.70 |
| Trunk fat percentage (TFP) | 7.48 (5.66–9.88) | <0.001 | 14.18 | 6.65 (5.04–8.78) | <0.001 | 13.38 |

Models assessed sex differences in OAC risk by adiposity trait. 656 cases of OAC were identified (15.9% in women). All Cox models were run with one adiposity trait at a time and additionally stratified by age at recruitment, deprivation index, region, and adjusted for qualifications, ethnicity, height, smoking status, alcohol, and prevalent gastro-oesophageal reflux disease.

*BMI adjustment results are shown for completeness; cancer-specific multivariable BMI results were presented in Chapter 4. Here, they were recalculated owing to a slight reduction in total cancer cases with additional cohort exclusions. The Chapter 4 HR from a minimally adjusted Cox model (stratified by age at recruitment, Townsend deprivation index, and region, and no further adjustments) was 6.01 (95% CI 4.90–7.37; p<0.001).

The percentage change in the natural logarithm of the hazard ratios is additionally given in Appendix Table 5B.

**Table 5.9 Hazard ratios (95% confidence intervals) comparing men with women for risk of kidney cancer [ICD-10 C64-C65], adjusted for one adiposity trait at a time.**

| Adiposity Trait | Continuous adjustment | | | Categorical adjustment | | |
|---|---|---|---|---|---|---|
| | HR (95% CI) | p-value | Z-score | HR (95% CI) | p-value | Z-score |
| *No adiposity trait adjustment* | *1.82 (1.54-2.15)* | *<0.001* | *7.17* | *1.82 (1.54-2.15)* | *<0.001* | *7.17* |
| BMI* (kg/m$^2$) | 1.75 (1.49–2.07) | <0.001 | 6.65 | 1.74 (1.47–2.05) | <0.001 | 6.51 |
| Body fat percentage (BFP) | 2.49 (2.04–3.03) | <0.001 | 9.03 | 2.43 (1.99–2.97) | <0.001 | 8.71 |
| Hip circumference (HC; cm) | 1.97 (1.67–2.33) | <0.001 | 7.96 | 1.95 (1.65–2.31) | <0.001 | 7.84 |
| Waist circumference (WC; cm) | 1.57 (1.33–1.86) | <0.001 | 5.27 | 1.58 (1.33–1.88) | <0.001 | 5.27 |
| Waist-to-hip ratio (WHR) | 1.26 (1.05–1.52) | 0.015 | 2.42 | 1.36 (1.13–1.65) | 0.001 | 3.17 |
| Waist-to-height ratio (WHtR) | 1.57 (1.33–1.86) | <0.001 | 5.25 | 1.59 (1.34–1.88) | <0.001 | 5.33 |
| Trunk fat percentage (TFP) | 2.12 (1.78–2.53) | <0.001 | 8.42 | 2.11 (1.77–2.52) | <0.001 | 8.38 |

Models assessed sex differences in kidney cancer risk by adiposity trait. 1349 cases of kidney cancer were identified (35.7% in women). All Cox models were run with one adiposity trait at a time, and additionally stratified by age at recruitment, Townsend deprivation index, and region and adjusted for education (qualifications), ethnicity, height, smoking status, alcohol, and prevalent hypertension.

*BMI adjustment results are shown for completeness; cancer-specific multivariable BMI results were presented in Chapter 4. Here, they were recalculated owing to a slight reduction in total cancer cases with additional cohort exclusions. The Chapter 4 HR from a minimally adjusted Cox model (stratified by age at recruitment, Townsend deprivation index, and region, and no further adjustments) was 2.11 (95% CI 1.89-2.36; p<0.001).

The percentage change in the natural logarithm of the hazard ratios is additionally given in Appendix Table 5C.

## 5.4 Discussion

The analyses in this chapter were undertaken to examine whether sex differences in adiposity measures beyond body mass index (BMI) may provide insights into the male predominance of oesophageal adenocarcinoma, and cancers of the rectum and kidney. Six adiposity measures including BMI (retained as a reference) were selected *a priori* and examined: body fat percentage (BFP), waist circumference (WC), hip circumference (HC), waist-to-hip ratio (WHR), waist-to-height ratio (WHtR), and trunk fat percentage (TFP).

For oesophageal adenocarcinoma, men had a substantially higher incidence than women, with an approximately five-fold increased risk when models were adjusted for BMI and cancer risk factors. Adjustment for measures of central adiposity instead of BMI—waist circumference and waist-to-hip ratio—partially attenuated this association, reducing the relative risk to approximately 4.4-fold and 3.5-fold higher in men, respectively. Despite observing an attenuation, a marked male predominance remained, and further work understanding sex differences in oesophageal adenocarcinoma is needed.

For rectal cancer, men had 1.7-fold higher incidence compared to women, and it was observed that neither adjusting for BMI nor central-adiposity traits materially altered the gap. It is likely the differences may not be attributable to differential adiposity levels in men and women, and perhaps other risk factors (hormonal and health-related) may be relevant (described in Chapter 4, Section 4.4.4).

For kidney cancer, men had a 1.7-fold higher incidence compared to women when models were adjusted for BMI and cancer risk factors. Adjustment for measures of central adiposity instead of BMI—waist circumference and waist-to-hip ratio—strongly attenuated the association closer to 1.36-fold and 1.58-fold higher in men, respectively. It may be inferred central adiposity contributes strongly towards the sex differences of kidney cancer risk.

Excess adiposity is a well-established causal risk factor for both oesophageal adenocarcinoma and kidney cancer risk (Keum *et al.*, 2017, Lauby-Secretan *et al.*, 2016). The findings from our sex-stratified analyses of adiposity traits and cancer risk in men and women are broadly consistent with other prospective cohort studies of oesophageal adenocarcinoma (Sanikini *et al.*, 2020a, Sanikini *et al.*, 2020b), kidney cancer (Pischon *et al.*, 2006), and pan-cancer studies that included these cancer endpoints (Parra-Soto *et al.*, 2021, Rask-Andersen *et al.*, 2023).

This analysis benefits from notable strengths. Leveraging the large-scale UK Biobank cohort, sex differences in cancer risk across a comprehensive set of adiposity traits were investigated, including both anthropometric and bioimpedance-derived measures. The large sample size enabled sex-stratified analyses with sufficient power for kidney and rectal cancer, and for oesophageal adenocarcinoma (OAC). Despite lower case numbers for oesophageal adenocarcinoma in women, this study was relatively well powered compared to previous prospective evaluations in the UK Biobank (Sanikini *et al.*, 2020a). Additionally, adiposity measurements were assessed by trained research clinic staff instead of being self-reported, and data on high-quality body composition was available (i.e. DXA- and MRI-derived adiposity measurements) in a subsample, which allowed for assessing the associations of commonly used adiposity measurements with 'gold standard' measurements.

However, some limitations must be acknowledged. First, the interpretation of contributions from waist-to-hip ratio warrants caution and further investigations. As an adiposity trait, it is relatively independent of BMI and may be better representative of the sex differences in adiposity distributions, however, as a ratio it may obscure whether risk attenuation is driven by waist or hip circumference. A high waist-to-hip ratio may indicate increased abdominal obesity and/or reduced gluteal muscle mass (since hip circumference reflects not only fat mass but also gluteal muscle mass and bone structure) (Keum *et al.*, 2017). In addition, waist-

to-hip ratio was slightly more strongly correlated with BMI in men compared to women, therefore, a possibility of unwanted adjustment for sex itself remained. Second, as adiposity was assessed only at baseline, they are subject to potential regression dilution bias, whereby within-person variability and measurement error over time may attenuate observed associations. They may be particularly relevant when comparing adiposity traits with differences in long-term reproducibility. Consequently, the relative extent of attenuation observed after adjustment for different adiposity measures may have been influenced by differences in measurement stability over follow-up. Third, the evaluated adiposity traits do not distinguish between visceral and subcutaneous adipose tissue, which are implicated to have distinct influence on carcinogenesis (Keum *et al.*, 2017). Fourth, as multiple correlated adiposity traits were examined, the possibility of inflated type I error due to multiple testing should be considered when interpreting the findings. The analyses were hypothesis-driven and focused primarily on the consistency and attenuation of associations across biologically related adiposity traits. More conservative inferential approaches, including the use of 99% confidence intervals or formal correction for multiple testing, may nevertheless be informative in future analyses.

In these prospective analyses evaluating contributions from adiposity traits towards sex differences, it was observed adiposity traits (waist circumference and waist-to-hip ratio) strongly contributed to the observed male predominance of kidney cancer and partially contributed to the male predominance of oesophageal cancer and were not associated with risk of rectal cancer. These findings may be a valuable reference for future investigations of sex differences in cancer risk for these cancers.

# Chapter 6 Lung adenocarcinoma in Never Smokers: Investigating Hypotheses for Increased Risk in Women

## 6.1 Introduction

Chapter 6 details the epidemiology of lung adenocarcinoma (LUAD) in never smokers, highlights the persistent female predominance in incidence that is not fully explained by smoking behaviour, and reviews established and emerging risk factors relevant to never-smoker LUAD that may contribute to the female predominance. Findings from Chapter 4 demonstrated that the higher risk of LUAD among women persisted after adjustment for smoking and when analyses were restricted to never smokers (additionally adjusting for second-hand smoke and residential air pollution PM 2.5 levels) (Table 6.1). Here, three risk factors are additionally considered: household air pollution from cooking and heating fuel use (an established risk factor for lung cancer), and two candidate risk factors—prevalent asthma (a risk factor with probable evidence for lung adenocarcinoma in never smokers) and menopausal age as proxies for sex differences in hormonal exposures. The selection of these candidate risk factors was informed by three considerations: (1) epidemiological evidence suggesting an association with LUAD or lung cancer risk in never smokers, (2) evidence hypothesizing these risk factors may contribute to sex-specific differences in LUAD risk, and (3) the availability of data for risk factors within the UK Biobank cohort. Together, these considerations provide the conceptual framework that guided the analyses presented in this chapter.

**Table 6.1 Sex differences in risk of lung adenocarcinoma for men compared to women (results from Chapter 4; Table 4.2)**

| Model[1] | Cases (Women/Men) | Hazard Ratio (95% CIs) | P-value |
|---|---|---|---|
| Analyses performed in the main cohort (**470,720 participants**) | | | |
| Minimally adjusted | 825/697 | 0.98 (0.89-1.09) | 0.251 |
| Multivariable adjusted | 825/697 | 0.72 (0.62-0.84) | <0.0001 |
| Analyses performed in the **never smokers cohort** (**257,390 participants**)[2] | | | |
| Minimally adjusted | 174/80 | 0.69 (0.53-0.91) | 0.008 |
| Multivariable adjusted[3] | 174/80 | 0.62 (0.42-0.93) | 0.002 |

**Hazard ratios and 95% confidence intervals (CI) comparing risk of cancer in men to women for lung adenocarcinoma.**

[1]Cox proportional hazards (PH) regression models with attained age as the underlying time scale were used to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) for the risk of cancer in men compared to women. Minimally adjusted models were run stratifying for age, fifths of Townsend deprivation index and region. Multivariable adjusted models additionally adjusted for ethnicity, qualifications, height, BMI, smoking (status and intensity), and alcohol. Attenuation from minimally to multivariable adjusted models was linked to smoking, indicating adjustment for smoking revealed the higher risk for women.

[2]Analyses repeated in a cohort restricted to never smokers (participants who reported having smoked less than 100 cigarettes in their lifetime).

[3]Models were additionally adjusted for exposure to second-hand smoke (at home and outside the home) and ambient air pollution levels (pm2.5).

## 6.1.1 Epidemiology

Lung cancer is the third most commonly diagnosed cancer worldwide and the second most common cancer in both men and women in the UK (Cancer Research UK, 2025, Murphy *et al.*, 2025). In the UK, approximately 71% of lung cancer cases are attributable to current or former tobacco smoking, and 1% to environmental tobacco smoke (Brown *et al.*, 2018, Cancer Research UK, 2025). Since the 1990s, lung cancer incidence in the UK has declined overall by roughly 8% (Cancer Research UK, 2025). However, this overall trend diverges markedly by sex:

incidence rates have increased by about one-third in women while decreasing by one-third in men (Cancer Research UK, 2025). These contrasting patterns largely reflect sex and cohort differences in tobacco exposure, as women adopted smoking more gradually during the 1950s than men, and the decline in smoking prevalence has been slower and more delayed among women (Brown *et al.*, 2018, Peters *et al.*, 2014, Peto *et al.*, 1976). In 2024, 12.2% of men versus 8.7% of women reported having ever smoked in England, and these estimates have remained similar since 2011 (Office for Health Improvement & Disparities, 2024).

In parallel with the decline in tobacco-related lung cancer, there has been a growing recognition of lung cancer in never smokers as a distinct public health concern, with a disproportionate rise in incidence observed among women—particularly at younger ages (Jemal *et al.*, 2018, Jemal *et al.*, 2025, Murphy *et al.*, 2025). An individual who has smoked fewer than 100 cigarettes in their lifetime is generally considered a 'never smoker'. If lung cancer in never smokers were considered as distinct, it would rank as the seventh most common cancer worldwide and the eighth leading cause of cancer-related death in the UK (Cancer Research UK, 2025, Murphy *et al.*, 2025). Lung adenocarcinoma (LUAD) is a histological subtype of non-small cell lung cancer originating in the glandular tissue of the peripheral lung, and is the predominant subtype among never smokers—accounting for approximately 60–80% of the cases (Murphy *et al.*, 2025). Notably, LUAD is the only subtype with both higher and increasing incidence in women compared to men, particularly amongst younger age groups (Jemal *et al.*, 2018, Jemal *et al.*, 2025, Murphy *et al.*, 2025).

### 6.1.2 Sex Differences in Incidence Beyond Smoking

While tobacco smoking remains the dominant cause of lung cancer overall, including lung adenocarcinoma (LUAD), several studies suggest the higher incidence of LUAD in women is not fully explained by sex differences in smoking behaviour (Jemal *et al.*, 2018, Jemal *et al.*, 2025, Lan *et al.*, 2012). Although smoking prevalence amongst women has become recently

similar to men, it does not exceed it (Jemal *et al.*, 2018, Office for Health Improvement & Disparities, 2024, Peters *et al.*, 2014). The use of other tobacco products—such as cigars and smokeless tobacco (chewing tobacco)—is also lower among women than men (Jemal *et al.*, 2018, Nelson *et al.*, 2006). A GWAS of lung cancer susceptibility loci amongst women of Asian ancestry observed no association for lung cancer at 15q25 in never smoking women, which is a locus associated with lung cancer, smoking behaviour, and nicotine addiction (Lan *et al.*, 2012, The Tobacco and Genetics Consortium, 2010).

Next, it has been proposed that differences in lung cancer histological subtypes and in the pace of risk reduction following smoking cessation may partly underlie the sex differences in incidence (Bain *et al.*, 2004, Jemal *et al.*, 2018). LUAD, specifically, has a slower decline in risk reduction after smoking cessation compared with the other histological subtypes such as squamous cell carcinoma (Bain *et al.*, 2004, Jemal *et al.*, 2018, Kenfield *et al.*, 2008). A study in the Nurses' Health cohort reported annual risk reduction of lung adenocarcinoma was 8% compared to 17% for squamous cell carcinoma post smoking cessation (Kenfield *et al.*, 2008). However, direct prospective evidence amongst former smokers stratified by sex and histology is limited, and thus it has not been evaluated in the context of understanding sex differences. Previously it has also been hypothesized that the deleterious effects of tobacco smoking may be more severe in women than in men, however, prospective cohort studies have not reported any sex differences (BMJ Research News, 2013, Jemal *et al.*, 2018, Peto *et al.*, 2000).

Lastly, the analyses of sex differences in LUAD risk from Chapter 4 noted a higher risk for women even after adjusting for extrinsic and intrinsic risk factors (men-to-women HR 0.72; 95% CI 0.62-0.84, $p<0.0001$) and when assessing sex differences in a cohort of never smokers (HR 0.62, 95% CI 0.42-0.93, $p=0.002$) (Table 6.1). Collectively, these patterns indicate that women generally have lower cumulative tobacco exposure, yet they experience higher LUAD

incidence, and perhaps other risk factors beyond smoking may contribute to this sex disparity.

### 6.1.3 Established LUAD Risk Factors in Never Smokers

The well-established risk factors for LUAD in never smokers are involuntary exposure to tobacco smoke, occupational carcinogenic exposure (section 2.4.9 lists exposures in detail), radon, outdoor air pollution (particularly fine particular matter; $PM_{2.5}$), indoor air pollution (particularly, exposure to combustion fumes from cooking and heating fuels in poorly ventilated houses), history of lung disease, history of chest radiotherapy, family history of lung cancer, Asian ancestry, and germline variants (Figure 6.1) (Centre for Disease Control and Prevention, 2024, Kratzer *et al.*, 2024, Malhotra *et al.*, 2018, Murphy *et al.*, 2025). History of lung diseases are classified as host susceptibility risk factors for increasing lung cancer risk and thought to influence cancer risk through chronic inflammation and lung damage (Kratzer *et al.*, 2024, Thun *et al.*, 2017). They include chronic bronchitis, chronic obstructive pulmonary disease, and tuberculosis infection, which are associated with cancer risk regardless of smoking status, and asthma and pneumonia, associated with risk predominantly among never smokers (Kratzer *et al.*, 2024, Pirie *et al.*, 2016, Thun *et al.*, 2017). At the molecular level, lung cancer in never smokers exhibits distinct biological characteristics compared to lung cancer in smokers (Zhang *et al.*, 2021). Besides being predominantly adenocarcinoma histological subtypes, they display a relatively low mutational burden, have longer latency (~9 years on average), and higher frequencies of EGFR and KRAS mutations and ALK gene rearrangements, as well as germline variants in AR and CYP21A2, suggestive of underlying hormonal or genetic predispositions (Murphy *et al.*, 2025, Zhang *et al.*, 2021). In addition to these, several risk factors—including chronic inflammation, adult height, hormonal influences, and viral infections such as Epstein-Barr virus and pulmonary tuberculosis—have been implicated to be associated with LUAD (Jemal *et al.*,

2025, Murphy *et al.*, 2025, Pirie *et al.*, 2016, Ragavan and Patel, 2022). Finally, while lung cancer in never smokers may be investigated as a separate entity from lung cancer in smokers, studies acknowledge that there may be some inaccuracies on self-reported smoking status and it may be slightly influenced by social desirability biases (Jemal *et al.*, 2018, Kukhareva *et al.*, 2022, Murphy *et al.*, 2025).

### 6.1.4 Relevance of Risk Factors towards Sex Differences

Some of the risk factors described above have been hypothesized to contribute to the greater risk observed among women, but require further study (Jemal *et al.*, 2018, Jemal *et al.*, 2025, Murphy *et al.*, 2025, Ragavan and Patel, 2022, Zhang *et al.*, 2021). Here, research on the risk factors is reviewed and categorized based on the criteria defined in Chapter 1. Risk factors may be considered extrinsic (environmental and health-related factors which may have sex differences in their distribution) and intrinsic (biological and sex-related factors that may modify risk independently or in interaction with environmental carcinogens) (Figure 1.2).

Among extrinsic risk factors, sex differences in second-hand smoke exposure were considered important, however, adjustment for it in Chapter 4 analyses (Table 6.1) did not materially alter the hazard ratio for sex differences in LUAD amongst never smokers. This was in line with what was expected, as second-hand smoke exposure in non-smoking adults was similar between the sexes in the cohort, and has also been reported to be generally similar by sex in the UK (in 2019 it was reported 28% men and 24% women were exposed to second-hand smoke) (NHS Digital, 2019).

Second, ambient air pollution, notably PM2.5, has been associated with increased lung cancer risk in large-scale prospective cohort studies (Loomis *et al.*, 2014, Swanton *et al.*, 2022). PM2.5 represents particulate matter ≤2.5 μm in aerodynamic diameter, comprising organic and inorganic particles from sources such as vehicle emissions, industrial combustion, and domestic fuel use. Ambient PM2.5 is not expected to differ meaningfully by

sex within the same residential area and was therefore adjusted for in the Chapter 4 never-smoker model to address residual confounding from a well-established lung cancer risk factor rather than as a candidate explanation for the observed sex difference (Chapter 4, Section 4.2.2). However, its inclusion did not materially alter the effect estimates for any lung cancer subtype, including LUAD. PM2.5 was selected over other ambient pollutants such as $NO_2$ on the basis of its stronger and more consistent epidemiological evidence as a contributor to lung cancer risk in never smokers (Hamra *et al.*, 2014, Hvidtfeldt *et al.*, 2021, International Agency for Research on Cancer, 2013, Puett *et al.*, 2014). One recent prospective cohort study reported an independent $NO_2$ and LUAD association after adjustment for PM2.5, with no evidence of effect modification by sex (Fisher *et al.*, 2025). However, $NO_2$ has only weak direct mutagenic effects and is generally interpreted as a proxy for other carcinogenic components of the traffic-related pollution mixture rather than as an independent cause (Fisher *et al.*, 2025, Koehler *et al.*, 2013). Additionally, both PM2.5 and $NO_2$ are also susceptible to exposure misclassification, since modelled estimates assigned to a baseline residential address do not capture residential mobility, indoor or occupational exposures, or temporal variation across follow-up (Fisher *et al.*, 2025, Huang *et al.*, 2021, Hvidtfeldt *et al.*, 2021). Future analyses could usefully incorporate more detailed air pollution measures and finer-grained exposure histories. Some studies have hypothesised that women may be more susceptible to ambient air pollution either directly or via interaction with hormonal risk factors, though robust evidence is lacking (Fisher *et al.*, 2025, Huang *et al.*, 2021, Hvidtfeldt *et al.*, 2021).

In contrast, household air pollution represents a more plausible source of sex-differences in exposure (Figure 6.1). Women in home settings may be exposed to higher levels of indoor air pollution from combustion of cooking and heating fuels, especially in spaces with poor ventilation (Hosgood *et al.*, 2015, Malhotra *et al.*, 2018). The IARC classifies emissions from the burning of solid fuels (for example coal) for domestic purposes as carcinogenic, based

largely on evidence from studies in China (Loomis *et al.*, 2014, Malhotra *et al.*, 2018). Systematic reviews provide estimates of the increased risk of lung cancer associated with exposure to be twofold, and higher risk in women (around twofold compared to men) reflecting their higher levels of exposure (Bruce *et al.*, 2013). There is however geographical variation in reported risk estimates, and they are highest in southern China (> 2.5) (Bruce *et al.*, 2013). A positive but relatively modest association between indicators of air pollution and lung cancer risk has also been reported in populations exposed to less severe conditions than women in Asian countries, for example populations in central and eastern Europe and other regions (Hosgood *et al.*, 2010, Lissowska *et al.*, 2005). The IARC International Multicentre Case-Control Study in Eastern and Central Europe and the UK of 2,681 cases and 3,118 controls, assessed the risk of lung cancer in relation to indoor air pollution. They reported the odds ratio of lung cancer associated with solid fuel use was 1.22 (95% CI, 1.04-1.44) for cooking or heating, 1.37 (0.90-2.09) for solid fuel only for cooking, and 1.24 (95% CI: 1.05-1.47) for solid fuels used for both cooking and heating (Lissowska *et al.*, 2005). An analysis with a larger sample size in the international lung cancer consortium of 5,105 cases and 6,535 controls reported odds ratios for predominant coal users were 1.64 (95% CI, 1.49-1.81), particularly coal users in Asia were 4.93 (3.73-6.52), and predominant wood users in North American and European countries were 1.21 (1.06-1.38) for risk of lung cancer, and that estimates were similar for never-smoking women (Hosgood *et al.*, 2010).

Other environmental carcinogens, such as occupational exposures (e.g. asbestos, cadmium, diesel exhaust) are less likely to contribute to the female predominance observed in never smokers. Historically, these exposures have been more prevalent among men, and their decline over recent decades may partly explain the greater reduction of lung cancer incidence in men compared with women (Jemal *et al.*, 2025). Radon exposure at home has been considered a plausible contributor (as women generally spend more time at home), however a case-control study of 7148 cases of lung cancer and 14 208 controls reported lung-cancer

risk increased by 8.4% (95% CI 3.0%–15.8%; $P = 0.0007$) per 100 Bq/m$^3$ increase in measured radon, and no significant heterogeneity for sex was detected (Darby *et al.*, 2005). Infectious agents such as TB and EBV are also less likely to contribute as the prevalence of exposures is generally higher in men compared to women (Jemal *et al.*, 2025).

Intrinsic factors may also contribute to sex disparities. Chronic respiratory conditions, such as asthma, which have a higher prevalence in women in adulthood, may predispose to lung damage and persistent inflammation, potentially increasing LUAD risk (Kole, 2024 (Kole *et al.*, 2024)). A prospective investigation of lung cancer in never smokers using data from the Million Women Study reported an association between asthma and lung cancer risk among never smokers (1.32, 1.10–1.58, $p = 0.003$) (Pirie *et al.*, 2016).

Hormonal factors have also been implicated, as most LUAD incidence in women has been observed in younger women (Jemal *et al.*, 2018, Jemal *et al.*, 2025). However, current prospective evidence is inconsistent in supporting whether sex-differences in circulating sex hormones or reproductive factors (proxies for sex hormone levels) may be associated with risk for women (Denos *et al.*, 2024, Elbasheer, 2024, Pirie *et al.*, 2016, Wilunda *et al.*, 2021). An analysis in the MWS found no significant associations with hormonal factors and cancer risk (Pirie *et al.*, 2016). A prospective analysis in the China Kadoorie Biobank of female never-smokers reported ever oral contraceptive use was associated with higher risk of lung cancer (HR 1.16; 95% CI 1.02-1.33), and longer average breastfeeding was associated with decreased risk (0.86, 0.78-0.95) (Elbasheer *et al.*, 2024).

### 6.1.5 Research Objective

Following the review of epidemiological evidence for risk factors associated with lung cancer in never smokers or LUAD, the evidence hypothesizing these risk factors may contribute to sex-specific differences in LUAD risk, and the availability of data for risk factors within the UK Biobank, three risk factors were selected for further analyses for sex differences of lung

adenocarcinoma risk in never smokers: household air pollution from cooking and heating fuel use (an established risk factor for lung cancer), and two candidate risk factors—prevalent asthma (a risk factor with probable evidence for lung adenocarcinoma in never smokers) and menopausal age as proxies for sex differences in hormonal exposures (Figure 6.1).

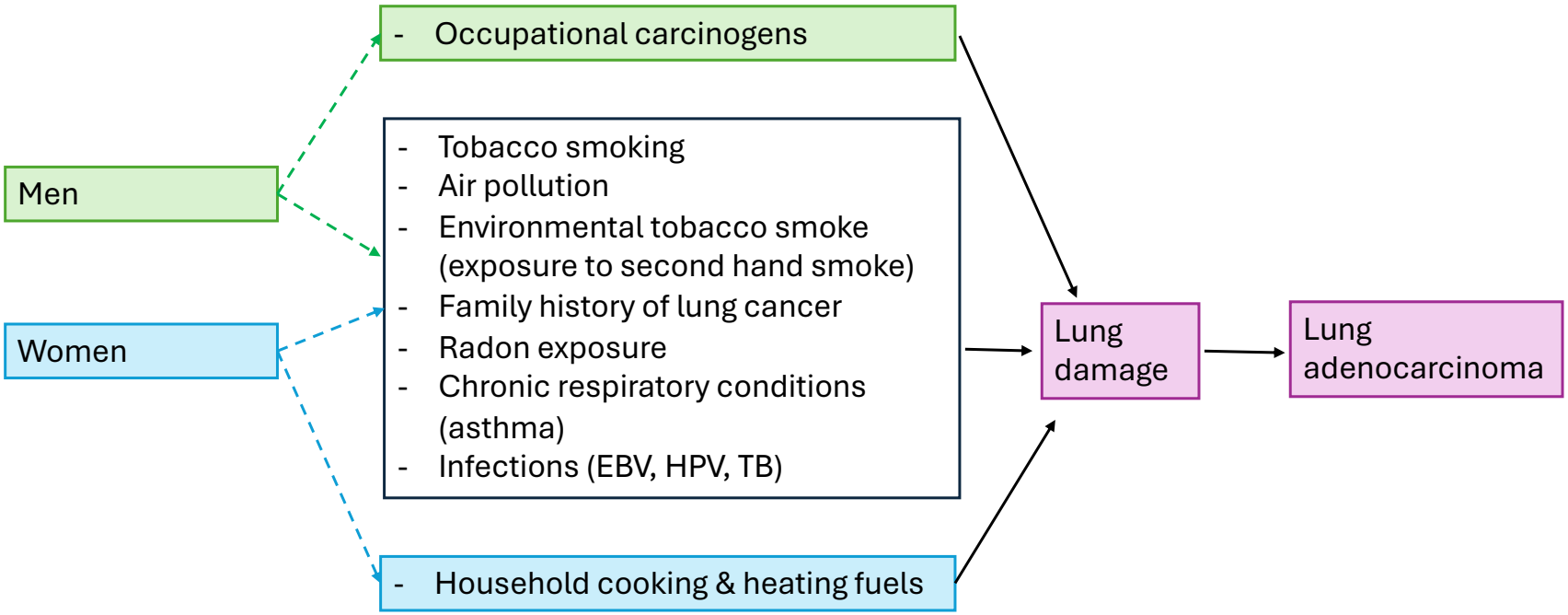



**Figure 6.1 Conceptual directed acyclic graph (DAG) illustrating how sex-related differences in the distributions of established risk factors may contribute to observed differences in lung adenocarcinoma (LUAD) incidence among never smokers.**

Risk factors included here are established component causes that may contribute probabilistically to lung damage and subsequent LUAD development. Solid arrows indicate hypothesised aetiologic relationships between exposures, intermediate lung damage, and LUAD, without implying a single necessary or sufficient causal pathway. Dotted arrows denote sex differences in the prevalence, intensity, or timing of exposures, which may contribute to the observed sex differences in incidence of LUAD.

## 6.2 Methods

Sex-stratified Cox proportional hazards models were used to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) for associations of household fuel use and asthma with lung adenocarcinoma (LUAD) risk among never smokers. Baseline multivariable-adjusted models were stratified by age at recruitment, Townsend deprivation index, and region, and additionally adjusted for ethnicity, highest educational qualification, height, and alcohol consumption. Further details of the Cox modelling approach and variable coding are provided in Chapter 3.

The objective was to assess whether adjustment for selected covariates (household fuel use and prevalent asthma) would attenuate the association between sex and LUAD risk; however, as neither covariate was materially associated with LUAD risk in sex-stratified analyses, these covariates were not further examined in analyses of sex differences.

Household cooking and heating fuel use was ascertained at recruitment as part of the touchscreen questionnaire participants filled out (data field 6139). One question on gas or solid-fuel used for cooking/heating was 'do you have any of the following in your home? (you can select more than one answer)'. Possible responses were: 'gas/hob cooker', 'gas fire used regularly in winter', 'open solid-fuel fire used regularly in winter', 'none of the above', 'do not know', and 'prefer not to answer'. A mutually exclusive four-level exposure category for was derived from the three baseline array slots (instance 0, arrays 0-2). Participants who selected 'none of the above' were categorised as 'electric' users only and used as the reference category in analyses. Participants who selected 'gas hob/cooker' and/or 'gas fire used regularly in the winter' in the absence of solid-fuel fire were categorised as 'gas' users only. Participants who selected both gas and solid fuel use options were categorised as 'solid and gas' users. Participants who selected 'open solid-fuel fire' were categorised as 'solid-fuel' users only. Participants who selected 'do not know' or 'prefer not to answer' were coded as

'missing'. Categorisation was implemented hierarchically, beginning with the electric-only category, followed by gas-only, combined solid-fuel and gas, and solid-fuel-only categories.

Prevalent asthma was ascertained in the cohort using ICD-10 code J45, or through self-reports during the interview portion of the participant recruitment process conducted by a medical professional at the recruitment centre (data-field 20002; code 1111).

To explore whether sex differences in LUAD risk varied by menopausal age, analyses were conducted using age at cancer diagnosis as a proxy for menopausal status. LUAD cases were categorised according to age at diagnosis into two groups: <50 years, 50-55 years, and >55 years. A small number of cases were diagnosed with LUAD in the cohort before age 50 (n=3) and between ages 50-55 (n=7), therefore sex-specific analyses were not performed in these groups. Sex differences in LUAD risk were examined in a restricted analysis including participants who were diagnosed after age 55 years. For this analysis, follow-up time was left-truncated at age 55 years, and baseline multivariable-adjusted Cox proportional hazards models were fitted.

## 6.3 Results

The final analytical cohort included 257,390 UK Biobank participants who self-reported as never smokers at recruitment, including 150,996 (58.7%) women and 106,394 (41.3%) men (Table 6.2). During a mean 10.6 years of follow-up (SD 2.1), 254 cases of lung adenocarcinoma (LUAD) were identified, 174 (68.5%) in women and 80 (31.5%) in men. The mean age for LUAD diagnosis was 67.2 years (SD 6.7) overall, 67.1 years (SD 6.5) in women and 67.2 years (SD 7.2) in men. The median age of diagnosis was 68.3 years overall (67.4 years in women and 69.5 years in men). Overall, no associations were observed between household cooking and heating fuel types or prevalent asthma with lung adenocarcinoma risk for never-smoking men and women in the cohort.

Gas was the most common household fuel (189,503 participants), followed by electric (43,987); combined solid and gas and solid-only use were less common (14,274 and 6,736 participants, respectively), particularly among men (Table 6.3). In multivariable-adjusted, sex-stratified Cox proportional hazards models among never smokers, household cooking and heating fuel use was not associated with lung adenocarcinoma risk (Table 6.3). In women, compared with electric fuel use, hazard ratios were 1.07 (95% CI 0.71–1.60) for gas use, 1.13 (95% CI 0.53–2.40) for combined solid and gas use, and 1.75 (95% CI 0.79–3.87) for solid fuel use alone. In men (80 cases), similarly, no evidence of an association was observed for any fuel category compared with electric fuel use, with hazard ratios of 0.92 (95% CI 0.52–1.59) for gas use, 0.93 (95% CI 0.30–1.36) for combined solid and gas use, and 0.98 (95% CI 0.22–4.34) for solid fuel use alone.

The prevalence of asthma in the cohort was slightly higher in women (13.7%) compared with men (12.3%) (Table 6.4). Asthma was not associated with risk of LUAD in women or men (Table 6.5). In women (172 cases), the HR comparing those with prevalent asthma to those without was 0.97 (95% CI 0.60–1.57; p = 0.91). In men (80 cases), the corresponding HR was 1.65 (95% CI 0.89–3.06; p = 0.11).

Additionally, in analyses restricted to participants diagnosed after age 55 years (244 cases), sex differences in LUAD risk were calculated as HR 0.61 (95% CI 0.41-0.91, p=0.016). These results were consistent estimates observed in the main analyses (HR 0.62; 95% CI 0.42-0.93, p=0.002

**Table 6.2 Characteristics of 257,390 participants of the UK Biobank who self-identified as never smokers at study recruitment.**

| | Women (%) | Men (%) | Total (%) |
|---|---|---|---|
| **Cohort size** | 150,996 (58.7) | 106,394 (41.3) | 257,390 |
| **Follow-up for cancer incidence** | | | |
| Number of incident cancers | 9,881 (67.4%) | 4,778 (32.6) | 14,659 |
| Person-years of follow up, (100,000s) | 16.1 | 11.2 | 27.3 |
| Mean follow-up time [SD] | 10.6 [2.0] | 10.5 [2.2] | 10.6 [2.1] |
| Median follow-up time [IQR] | 10.9 [10.2-11.7] | 10.9 [10.0-11.7] | 10.95 [10.1-11.7] |
| **Age, years** | | | |
| Mean age at recruitment [SD] | 55.8 [8.1] | 55.3 [8.2] | 55.6 [8.2] |
| <45 | 16,887 (11.2) | 13,709 (12.9) | 30,596 (11.9) |
| 45-49 | 22,380 (14.8) | 16,691 (15.7) | 39,071 (15.2) |
| 50-54 | 24,925 (16.5) | 17,647 (16.6) | 42,572 (16.5) |
| 55-59 | 27,628 (18.3) | 18,778 (17.6) | 46,406 (18.0) |
| 60-64 | 34,021 (22.5) | 22,554 (21.2) | 56,575 (22.0) |
| >65 | 25,155 (16.7) | 17,015 (16.0) | 42,170 (16.4) |
| **Region** | | | |
| London | 20,352 (13.5) | 13,846 (13.0) | 34,198 (13.3) |
| North-West England | 22,486 (14.9) | 16,661 (15.7) | 39,147 (15.2) |
| North-Eastern England | 17,790 (11.8) | 12,681 (11.9) | 30,471 (11.8) |
| Yorkshire & the Humber | 22,508 (14.9) | 15,881 (14.9) | 38,389 (14.9) |
| West Midlands | 13,155 (8.7) | 9,840 (9.2) | 22,995 (8.9) |
| East Midlands | 10,504 (7.0) | 7,048 (6.6) | 17,552 (6.8) |
| South-East England | 13,882 (9.2) | 9,519 (8.9) | 23,401 (9.1) |
| South-West England | 13,045 (8.6) | 8,757 (8.2) | 21,802 (8.5) |
| Wales | 6,303 (4.2) | 4,410 (4.1) | 10,713 (4.2) |
| Scotland | 10,971 (7.3) | 7,751 (7.3) | 18,722 (7.3) |
| **Deprivation Index, in quintiles** | | | |
| Most affluent | 33,315 (22.1) | 23,729 (22.3) | 57,044 (22.2) |
| 2nd | 32,166 (21.3) | 22,596 (21.2) | 54,762 (21.3) |
| 3rd | 31,249 (20.7) | 21,602 (20.3) | 52,851 (20.5) |
| 4th | 29,152 (19.3) | 20,276 (19.1) | 49,428 (19.2) |
| Most deprived | 24,939 (16.5) | 18,044 (17.0) | 42,983 (16.7) |

| | | | |
|---|---|---|---|
| Unknown[a] | 175 (0.1) | 147 (0.1) | 322 (0.1) |
| **BMI, kg/m²** | | | |
| Mean BMI [SD] | 26.97 [5.18] | 27.50 [4.17] | 27.19 [4.80] |
| <25 | 60,978 (40.4) | 29,390 (27.6) | 90,368 (35.1) |
| 25- | 54,908 (36.4) | 52,856 (49.7) | 107,764 (41.9) |
| 30- | 23,368 (15.5) | 18,780 (17.7) | 42,148 (16.4) |
| 35- | 11,742 (7.8) | 5,368 (5.0) | 17,110 (6.6) |
| **Weight, kg** | | | |
| Mean weight [SD] | 70.98 [13.95] | 85.07 [14.07] | 76.81 [15.62] |
| **Height, cm** | | | |
| Mean height [SD] | 162.31 [6.35] | 175.85 [6.90] | 167.91 [9.37] |
| **Qualifications** | | | |
| Prof Q/NVQ/HND/HNC/Degree or other professional qualification | 88,799 (58.8) | 72,510 (68.2) | 161,309 (62.7) |
| A levels | 8,901 (5.9) | 5,420 (5.1) | 14,321 (5.6) |
| O level/GSE/CSE | 28,954 (19.2) | 13,425 (12.6) | 42,379 (16.5) |
| None of the above | 21,691 (14.3) | 13,197 (12.4) | 34,888 (13.6) |
| Unknown[a] | 2,651 (1.8) | 1,842 (1.7) | 4,493 (1.7) |
| **Ethnicity** | | | |
| White | 139,818 (92.6) | 99,135 (93.2) | 238,953 (92.8) |
| Mixed Race | 879 (0.6) | 474 (0.4) | 1,353 (0.5) |
| Asian or Asian British | 4,782 (3.2) | 3,524 (3.3) | 8,306 (3.2) |
| Black or Black British | 3,393 (2.2) | 1,914 (1.8) | 5,307 (2.1) |
| Other | 1,710 (1.1) | 912 (0.9) | 2,622 (1.0) |
| Unknown[a] | 414 (0.3) | 435 (0.4) | 849 (0.3) |
| **Alcohol consumption, grams per day** | | | |
| <1g | 24,641 (16.3) | 7,761 (7.3) | 32,402 (12.6) |
| ≥1-<10 | 64,644 (42.8) | 28,453 (26.7) | 93,097 (36.2) |
| ≥10-<20 | 30,479 (20.2) | 25,456 (23.9) | 55,935 (21.7) |
| ≥20-<30g | 9,364 (6.2) | 17,010 (16.0) | 26,374 (10.2) |
| ≥30 | 4,169 (2.8) | 19,899 (18.7) | 24,068 (9.4) |
| Non drinkers | 16,811 (11.1) | 7,470 (7.0) | 24,281 (9.4) |
| Unknown[a] | 888 (0.6) | 345 (0.3) | 1,233 (0.5) |

[a] This category includes participants whose information was missing because they skipped the question, or they selected 'prefer not to answer'

**Table 6.3 Household fuel type categorisation in 257,390 never smokers in the UK Biobank.**

| Fuel type | Women | Men | Total |
|---|---|---|---|
| **Electric** | 25,649 (29) | 18,338 (17) | 43,987 (46) |
| **Gas** | 111,284 (126) | 78,219 (60) | 189,503 (186) |
| **Solid + Gas** | 8,548 (9) | 6,126 (1) | 14,274 (10) |
| **Solid** | 4,086 (8) | 2,640 (2) | 6,736 (10) |
| *Missing* | *1,419 (2)* | *1,071 (0)* | *2,490 (2)* |
| **Total** | 150,996 (174) | 106,394 (80) | 257,390 (254) |

Numbers are participants; lung adenocarcinoma cases shown in parentheses

**Table 6.4 Hazard ratios (HRs) and 95% confidence intervals (CI) for associations between household fuel type and lung adenocarcinoma [C34] risk, in 149,577 women and 105,223 men from the UK Biobank.**

| | **Women** ***172 cases*** | | | **Men** ***80 cases*** | | |
|---|---|---|---|---|---|---|
| | **HR (95% CI)** | **Z-score** | **p-value** | **HR (95% CI)** | **Z-score** | **p-value** |
| Electric | *(ref)* | .. | .. | *(ref)* | .. | .. |
| Gas only | 1.07 (0.71-1.60) | 0.30 | 0.76 | 0.92 (0.52 – 1.59) | -0.31 | 0.75 |
| Solid + Gas | 1.13 (0.53-2.40) | 0.31 | 0.75 | 0.93 (0.30 – 1.36) | -1.41 | 0.15 |
| Solid only | 1.75 (0.79-3.87) | 1.39 | 0.16 | 0.98 (0.22 – 4.34) | -0.02 | 0.98 |

Multivariable adjusted Cox PH models were stratified by age at recruitment, Townsend deprivation index, and region, and additionally adjusted for ethnicity, qualifications, BMI, height, and alcohol. Analyses were performed in a cohort of never smokers excluding 2,492 participants with missing information on household fuel use. Electric fuel use was the reference category.

**Table 6.5 Prevalent asthma in 257,390 never smokers in the UK Biobank.**

| | Women | Men | Total |
|---|---|---|---|
| **Asthma cases** | 18,231 (19) | 11,675 (12) | 29,906 (31) |
| **Controls** | 132,765 (155) | 94,719 (68) | 227,484 (223) |
| **Total** | 150,996 (174) | 106,394 (80) | 257,390 (254) |

Numbers are participants; lung adenocarcinoma cases shown in parentheses.

**Table 6.6 Hazard ratios (HRs) and 95% confidence intervals (CI) for associations between prevalent asthma and lung adenocarcinoma [C34] risk, in 150,996 women and 106,394 men from the UK Biobank.**

| | **Women** ***172 cases*** | | **Men** ***80 cases*** | |
|---|---|---|---|---|
| | HR (95% CI) | p-value | HR (95% CI) | p-value |
| **Asthma (yes/no)** | 0.97 (0.60-1.57) | 0.91 | 1.65 (0.89-3.06) | 0.11 |

Multivariable adjusted Cox PH models were stratified by age at recruitment, Townsend deprivation index, and region, and adjusted for ethnicity, qualifications, BMI, height, alcohol, and additionally prevalent asthma.

## 6.4 Discussion

This chapter investigated potential contributors to the consistently higher incidence of lung adenocarcinoma (LUAD) among never-smoking women compared with never-smoking men. Three risk factors were evaluated: household cooking and heating fuel exposure, asthma, and effect modification by menopausal status. Across all analyses, none of these factors were associated with LUAD incidence in either men or women. One important consideration in interpreting these findings is that the higher risk of LUAD observed in women was relatively modest and may be partly attributable to limited statistical power, as there were fewer LUAD cases among never smokers overall. Studies with larger numbers of cases are therefore needed to provide greater insight into potential sex differences in LUAD risk among never smokers.

Another limitation is that there may potential misclassification of household cooking and heating fuel exposure. The UK Biobank assessment captured current fuel use at baseline and may not adequately reflect early-life exposure or cumulative duration and intensity. In addition, although lung cancer incidence differs by sex among never smokers, survival and mortality tend to favour women (Saha *et al.*, 2024), raising the hypothesis that women may be diagnosed earlier due to more proactive health-seeking behaviours. However, mean age at diagnosis was similar between men and women in the present analyses. Consistent with this, a large retrospective cohort study of over 21,000 never smokers in South Korea found no sex difference in lung cancer risk among screened individuals (adjusted HR 0.90, 95% CI 0.62–1.26) (Kim *et al.*, 2025). Taiwan remains the only country with a lung cancer screening programme for never smokers, with similar uptake between men and women (Murphy *et al.*, 2025).

Overall, sex differences in lung adenocarcinoma among never smokers remain incompletely understood and warrant further investigation.

# Chapter 7 Discussion and Conclusions

## 7.1 Key Findings and Research Directions

The aim of this thesis was to conduct a comprehensive investigation of sex differences in cancer incidence and to assess the extent to which observed disparities persisted after accounting for established and measured risk factors. It was hypothesized that each cancer site with an observable sex difference in its incidence may be attributable to differences in exposure to extrinsic—environmental and health-related—cancer risk factors, intrinsic differences through biological and sex-related characteristics, or a combination of both. Sex differences in incidence of 15 cancers (and 13 subtypes) were investigated systematically by evaluating contributions of extrinsic and intrinsic cancer risk factors using data from 470,720 participants in the UK Biobank prospective cohort study. Analyses were stratified by age, deprivation index, and region. The considered extrinsic risk factors included socio-demographic factors (education, ethnicity) and health-related behaviours (smoking, alcohol consumption, physical activity) and intrinsic risk factors included biological characteristics (height, body mass index) for all cancer endpoints, and additionally site-specific risk factors. Cancer site-specific risk factors were selected based on three considerations: evidence as established risk factors, epidemiological evidence supporting their contribution to observed sex differences, and availability of relevant data within the UK Biobank cohort.

Four cancer sites—oesophageal adenocarcinoma, bladder, rectum, and lung adenocarcinoma—had marked sex differences in incidence that remained upon evaluation of contributions from risk factors and therefore merit further investigation. The observed sex differences for these cancer sites were consistent after accounting for site-specific risk factors, and across analyses

restricted to never smokers, light alcohol drinkers, or participants with normal BMI (Table 7.1 summarises the main findings for cancer sites with higher risk in men compared to women).

Men had a higher incidence of oesophageal adenocarcinoma (nearly five-fold) compared to women. Measures of central adiposity and gastro-oesophageal reflux disease were also evaluated for their contributions to observed sex disparity. Waist circumference and waist-to-hip ratio were partially contributory to attenuating the relative risk from five-fold higher to 4.4- and 3.5-fold higher, respectively; however, a large male predominance remained. Bladder cancer had more than threefold higher incidence in men compared to women (no bladder cancer-specific risk factors were evaluated). Rectal cancer had 1.7-fold higher incidence in men, which remained also upon consideration of diabetes, inflammatory bowel disease, processed meat intake, and central adiposity measures.

Lung adenocarcinoma risk was higher in women compared to men (approximately 1.3-fold) and remained consistently elevated among never smokers (nearly 1.4-fold). In unadjusted analyses, the relative risk between men and women was close to the null; however, adjustment for established risk factors—particularly smoking—strengthened the association away from the null. This may reflect higher smoking prevalence among men masking an underlying higher risk of lung adenocarcinoma in women (50.0% men versus 39.9% women self-reported as current or former smokers). While smoking is a major risk factor for lung adenocarcinoma, the adjustment for smoking revealed the higher risk for women and thus may not be fully explained by smoking behaviours alone. Consequently, sex differences for lung adenocarcinoma in never smokers were investigated. The higher risk in women remained upon evaluation of contributions from household heating and cooking fuel exposure, a well-established risk factor for lung cancer, and asthma and menopausal status, candidate risk factors hypothesised to contribute to the observed sex differences. Altogether, the sex difference observed for lung adenocarcinoma for

women was relatively modest compared to the sex differences for cancer sites with a male predominance.

In addition to the main thesis findings, men had a 1.8-fold higher incidence of kidney cancer than women, observed upon additionally considering hypertension. Subsequently, contributions from measures of central adiposity were considered. Waist circumference and waist-to-hip ratio strongly contributed to attenuating the relative risk to 1.6-fold and 1.4-fold higher. However, similar to the findings for oesophageal adenocarcinoma, interpretation of contributions from waist-to-hip ratio warrants caution and further investigations. As an adiposity trait, it is relatively independent of BMI and may be better representative of the sex differences in adiposity distributions, however, as a ratio it may obscure whether risk attenuation is driven by waist or hip circumference. In addition, waist-to-hip ratio was slightly more strongly correlated with BMI in men compared to women, therefore, a possibility of unwanted adjustment for sex itself remained.

Future research on elucidating the exposures underlying the persistent sex differences observed across oesophageal adenocarcinoma, bladder, rectum, and lung adenocarcinoma, may enhance our understanding of their aetiology. For oesophageal adenocarcinoma, further investigation on biological (metabolic and sex-specific) factors is crucial. Additionally, similarly for kidney cancer, understanding mechanisms underpinning the exposures mediating central adiposity related contributions to the excess risk in men are also essential. For bladder cancer, future work considering sex differences in occupational exposures and genetics (X and Y chromosome) are needed to better understand the substantial excess risk observed in men. Similarly, the higher incidence of rectal cancer in men highlights the need to identify rectal cancer–specific risk factors that may disproportionately affect men and may not explained by established behavioural or metabolic exposures. Finally, studies with higher numbers of cases are essential to more robustly characterise the sex differences in lung adenocarcinoma and elucidate the role of environmental and biological factors contributing to the modest but

consistent excess risk observed in women. Importantly, the attenuation of sex differences observed for other cancer endpoints following adjustment for established risk factors is also informative and of relevance to future research (refer to Appendix Tables 7A and 7B, and Figure 4.1).

## 7.2 Strengths and Research Contributions

The comprehensive and systematic evaluation of sex differences in cancer incidence undertaken in this doctoral work represents a novel contribution to cancer epidemiology. A key strength lies in the use of an *a priori* framework to guide the selection and investigation of major cancer risk factors hypothesised to contribute to observed sex differences. The study design focused on the most common cancer sites, thereby minimising outcome selection bias and ensuring consistency in risk estimation across analyses. Next, the analytical approach adopted may be applied to investigations of sex differences in cancer incidence in other large prospective cohort studies and of other chronic diseases. Collectively, the rationale and analytical approach provide a valuable reference for future studies of cancer aetiology.

The strengths of the analyses are further reinforced by the use of UK Biobank, one of the largest prospective cohort studies to date, with an average follow-up of over a decade. The study was well powered for most cancer endpoints, included detailed information on cancer histological subtypes, and captured cohort-wide data on the majority of well-established cancer risk factors evaluated. Additionally, it had detailed information on several emerging risk factors for cancer. A key highlight of the study includes large numbers of participants who identified as never smokers and had normal BMI, which enabled robust comparisons of sex differences and assessment of potential residual confounding from smoking and excess adiposity, currently the two largest contributors of cancer incidence.

## 7.3 Limitations

This doctoral work has certain limitations. First, the analyses are likely to be subject to some degree of unmeasured and residual confounding, including, for example imperfect measurement of smoking and alcohol consumption. Sensitivity analyses of low alcohol consumption were restricted to light drinkers because relatively few participants reported never drinking alcohol (64.0% of the cohort were light drinkers compared with 7.9% who were never drinkers). As a result, residual confounding by alcohol may persist, which is particularly relevant for cancers of the oral cavity and rectum. Sensitivity analyses restricted to participants with normal body mass index (BMI) may be slightly influenced by unmeasured changes in weight over follow-up. Some sex differences could not be evaluated because certain cancer sites had insufficient numbers of cases for subgroup analyses, notably cancers of the anus and gastric cardia.

Second, there were putative or established risk factors that were not included in models either because they were not measured at all or only measured for a subset of the cohort, or because they were traits with some but limited evidence of their associations with cancer sites. These included occupations, carcinogenic infections, comorbidities, and other biological and health-related traits (for example, metabolic syndrome, sedentary behaviour, inflammatory markers). Further research is warranted to assess the extent to which contributions from these risk factors might attenuate the observed sex differences for some cancers.

Third, it may also be useful to replicate these analyses in prospective cohort studies with more cases for some cancer endpoints, including cancers of the gastric cardia, anus, and gallbladder, which had statistically significant or nearly statistically significant increased risk for one sex but could not be explored further due to relatively low power. For cancers of the gastric cardia and anus, statistically significant sex differences were observed but could not be explored comprehensively due to limited statistical power, particularly in analyses restricted to never smokers and participants with normal BMI. Fourth, the generalisability of the work is limited as

the UKB cohort study comprises mostly white European participants and participants who are relatively healthier than the UK population, and it would be useful to replicate a similar analysis in other study populations. Fifth, in the UK Biobank, it is not currently possible to distinguish sex from gender due to the survey questions, but this represents an important avenue for future research in prospective cohorts where this distinction can be factored into analyses.

Sixth, sex differences in risk for some cancers by anatomical subsites were not explored, nor was information available on tumour stage or histological grade, and analyses stratified by these categories may be insightful. This limitation also underscores the broader challenge of selecting appropriate cancer endpoints. For example, sex differences were examined across three endpoints: colorectal cancer as a combined outcome, and colon and rectal cancers assessed separately. The sex differences were null for colon cancer overall, but previous studies report sex-specific differences by colon subsite; women reportedly have higher right-sided colon cancers compared to men (Mauvais-Jarvis *et al.*, 2020, Murphy *et al.*, 2011, White *et al.*, 2018). Future studies should therefore adopt a cautious and considered approach, balancing efforts to elucidate cancer aetiology by subtype against the risk of overinterpretation or inflated findings arising from extensive subgroup analyses.

Lastly, as highlighted in the thesis introduction, it is important to recognise that language surrounding sex in public health and epidemiological research should be used cautiously and with context-specific clarification. Studies should reserve classification of sex as an established risk factor for contexts in which intrinsic biological determinants of sex are well-established causal factors, and otherwise emphasise the identification of underlying biological, health-related, and environmental factors that contribute to observed sex differences in disease incidence.

## 7.4 Future Research Directions

Future research evaluating the contributions of risk factors that may contribute to sex differences for some cancers is warranted. Evaluating any promising avenues includes exploring the methodological question of how exposures may be studied and their constraints. Three are considered here: occupational and environmental exposures, sex chromosome-linked genetic mechanisms, and learning from gender-affirming therapy cohorts as partial natural experiments for investigating hormone-mediated pathways relevant to sex differences in disease.

Understanding the role of occupational exposures to sex differences represents an important but methodologically challenging area for future research. Occupational exposures are typically distributed across a small proportion of the population, are correlated with socioeconomic and behavioural factors, and are imperfectly captured in general-population cohorts. For example, national registry linkage studies in Nordic countries, exemplified by the Nordic Occupational Cancer Study (NOCCA), provide near-complete occupational histories and cancer outcomes for approximately 15 million individuals followed for up to 45 years; however, they lack detailed lifestyle and biological data needed to disentangle occupational from behavioural contributions (Kjaerheim *et al.*, 2018, Pukkala *et al.*, 2009). Alternatively, a number of general population job exposure matrices have been developed to assess occupational exposures and applied to cohort studies; however, they may be prone to exposure misclassification bias (Sadhra *et al.*, 2016). Next, targeted analyses of specific occupations are constrained by risk estimates resting on small numbers of exposed cases within individual occupational categories and that job titles alone are not sufficient to characterise exposures (Matteis *et al.*, 2016). Another persistent challenge across approaches is distinguishing the effects of occupational exposure itself from the lifestyle profile correlated with particular occupations (Travis *et al.*, 2016). A prospective analysis on shift work and breast cancer in three cohort studies observed no apparent association between occupation and breast cancer incidence; the authors further noted occupations may be

intertwined with socioeconomic and behavioural patterns making analyses prone to residual confounding (Travis *et al.*, 2016). For example, women who do shift work are also more likely to be obese, to smoke, to take medications that help them sleep, and to prefer evenings to mornings (Travis *et al.*, 2016). Going forward, advances in untargeted biomarker measurement increasingly permit direct quantification of environmental exposures in stored biospecimens (Mancini *et al.*, 2020). For example, endocrine-disrupting chemicals proposed to increase breast cancer risk, per- and polyfluoroalkyl substances (PFAS), have been assayed in serum samples from the French E3N cohort, part of the European Prospective Investigation into Cancer and Nutrition (EPIC) (Mancini *et al.*, 2020). As such assays become scalable, retrospective biomarker measurement in large biobanks may permit objective exposure assessment for environmental and occupational carcinogens that has not previously been feasible at population scale, and these data will be valuable for further investigations of currently unexplained sex differences in cancer risk.

Next, genetic mechanisms linked to the X and Y chromosomes are also difficult to investigate in conventional epidemiological designs but may contribute to currently unexplained sex differences, particularly for cancers with marked male excess such as bladder cancer. Most GWAS have historically excluded the X chromosome from analysis owing to the methodological complications of dosage compensation and sex-specific genetic architecture, and the Y chromosome has been excluded almost universally due to its repetitive structure and homology with X (Sun *et al.*, 2023, Wise *et al.*, 2013).

A study that evaluated every GWAS paper published from 2010 to 2011 and reported in the National Human Genome Research Institute (NHGRI) GWAS Catalogue found only 33% (242 of 742 papers) included the X chromosome in analyses (Wise *et al.*, 2013). They additionally found that of more than 2,800 genome-wide significant associations reported at the time, the Y chromosome had no associations and the X chromosome only 15 associations at nine loci,

compared with 120 associations at 33 loci on chromosome 7, which is similar in physical size to the X chromosome. The authors concluded that the X chromosome should not be excluded in GWAS, and X-specific methods should be applied where appropriate, and technical, statistical, and cultural barriers should be further addressed (Wise *et al.*, 2013). Sun *et al.* (2023) demonstrated that, a decade later, this exclusion had not been resolved and had broadened to both sex chromosomes and polygenic prediction. They found that among GWAS summary statistics reported in 2021, only 25% included X-chromosome results and only 3% included Y-chromosome results. By November 2022, only about 0.5% of associated SNPs in the NHGRI-EBI GWAS Catalogue were on the X chromosome, and only nine of 447,939 associations were on the Y chromosome. Among 430 sampled polygenic score files, none included sex-chromosomal SNPs (Sun *et al.*, 2023). With regards to exclusion of the X chromosome, the authors identified constraints focused on autosome-centric pipelines across variant calling, quality control, imputation, principal-component analysis, association testing, summary-statistic reporting, PRS construction and downstream methods. For the Y chromosome, the authors emphasised additional constraints including sparse array coverage, repetitive sequence, incomplete reference representation, large structural variation, alignment artefacts, and age-related mosaic loss of Y in leukocytes affecting data quality and analyses.

Addressing these constraints would enable further avenues directly relevant to understanding sex differences. Future analyses should incorporate sex-aware quality control, appropriate modelling of X-chromosome dosage, pseudoautosomal regions, X-inactivation and Y-chromosomal structural variation. Mendelian randomization (MR) studies using sex-specific instruments and including X-linked variants for exposures with established sex-specific genetic architecture (for example, adiposity traits) may be applied to evaluate if sex-specific chromosome-linked pathways contribute cancer endpoints with notable sex differences. The interpretation of MR studies, however, also requires consideration of constraints. These may arise

if genetic instruments differ in strength by sex, pleiotropy operates through multiple biological pathways, or where the exposure itself is socially patterned (such as adiposity or sex hormones). Furthermore, using sex-combined genetic instruments for sex-stratified outcome analyses may also bias estimates (Gao *et al.*, 2021). Separately, mosaic loss of the Y chromosome in peripheral blood, a common somatic mutation in older men and a marker linked to the incidence of lung, bladder and haematological cancers, is another avenue for further research (Forsberg *et al.*, 2014, Thompson *et al.*, 2019). It may be further evaluated as a time-varying exposure in prospective cohorts with linked genotyping and cancer outcomes, particularly bladder cancer in men.

A further complementary approach includes studies of individuals receiving gender-affirming hormone therapy (GAHT) as they offer a partial dissociation between chromosomal sex and sex hormone effects, providing biological insights into sex differences that may not be accessible through conventional epidemiological studies (Yalcinkaya *et al.*, 2025). Cohort studies focused on hormonal contributions to cancer incidence report lower prostate cancer incidence in trans women receiving long-term androgen deprivation and oestrogen treatment (standardised incidence ratio 0.20, 95% CI 0.08-0.42) relative to the reference male population (de Nie *et al.*, 2020). Separately, lower breast cancer incidence in trans men receiving testosterone relative to cisgender women has been reported despite the retention of breast tissue (de Blok *et al.*, 2019). Immune function differs by sex across the life course, and the differences are shaped by sex hormones, chromosomes, age, puberty, pregnancy, menopause, and exogenous hormone exposure (Yalcinkaya *et al.*, 2025). Generally, women tend to have stronger innate and adaptive immune responses compared to men, faster infection clearance, stronger antibody responses and better vaccine outcomes (Yalcinkaya *et al.*, 2025). Men, by contrast, tend to have weaker responses to many infections and higher infectious mortality, partly attributed to the immunosuppressive effects of testosterone (Yalcinkaya *et al.*, 2025). A longitudinal cohort

analysis of 23 trans men receiving testosterone therapy found that testosterone directly modulates a cross-regulated axis between type-I interferon and tumour necrosis factor signalling, attenuating interferon responses and potentiating inflammatory pathways within months of treatment initiation, and that these effects are mediated by androgen receptor signalling rather than by the loss of oestradiol (Lakshmikanth *et al.*, 2024). Such findings illustrate how hormonal exposure may shift immune regulation across sexes, with relevance to sex differences in immune-related aspects of cancer aetiology, as well as infectious, inflammatory and autoimmune diseases (Lakshmikanth *et al.*, 2024, Yalcinkaya *et al.*, 2025). Nonetheless, interpretation requires care and is limited by small sample sizes, heterogenous treatment regimens, differences in anatomy and chromosomes, behavioural factors, minority stress, healthcare access, and reliance on immune biomarkers rather than long-term disease outcomes (Lakshmikanth *et al.*, 2024, Lett *et al.*, 2022, Yalcinkaya *et al.*, 2025).

Additionally, there are several cancer-related and chronic-stress-related risk factors elevated in trans individuals, resulting from institutionalised stigma and discrimination driving health-harming behaviours and disengagement from preventive care, while some are potentially hormone-mediated (Brown *et al.*, 2023). A cross-sectional analysis using the UK Clinical Practice Research Datalink identified elevated prevalence of several cancer-related risk factors among transgender individuals between 1988 and 2020 (Brown *et al.*, 2023). Transmasculine individuals had elevated prevalence of obesity, smoking, dyslipidaemia and hepatitis C infection compared with cisgender women, and an elevated prevalence of obesity, current alcohol use, diabetes, and HIV infection compared with cisgender men. Furthermore, transfeminine individuals had elevated prevalence of smoking, dyslipidaemia, diabetes, and hepatitis B, hepatitis C, and HIV infections and a decreased prevalence of obesity compared with cisgender women. Compared with cisgender men, transfeminine individuals had elevated dyslipidaemia, hepatitis C infection, and HIV infections (Brown *et al.*, 2023).

These findings also highlight the broader challenge of building an evidence base on GAHT cohorts. Trans and gender-diverse individuals remain under-represented in epidemiological research and routine medical practice, partly reflecting the discrimination they encounter in healthcare settings (Brown *et al.*, 2023, Lett *et al.*, 2022). A 2021 survey of almost 700 trans and gender-diverse individuals in the UK reported that 70% had experienced transphobia in medical settings and 14% had been refused healthcare of any kind by a GP because they were transgender, with these experiences more common among non-binary individuals and among Black people and people of colour (TransActual UK, 2021). Improving the inclusion of trans and gender-diverse individuals in cohort studies and routine data linkage would address inequities in their healthcare and also contribute biological insight into questions about sex and gender that benefit understanding of health more broadly.

A study designed to address these questions would ideally use a large prospective cohort with linked cancer registry data, repeated occupational histories, job-exposure matrices, biospecimen-derived exposure biomarkers, genome-wide genotyping including the X and Y chromosomes, and information on both sex assigned at birth and gender identity. Existing cohorts have enabled important progress along each of these avenues. The UK Biobank has supported the characterisation of mosaic Y loss and of sex-specific genetic architecture; however, its middle-aged-and-older recruitment window restricts the assessment of exposures established earlier in life, and its survey design does not include a comprehensive assessment of gender. Emerging biobanks such as Our Future Health in the UK and the All of Us Research Program in the US have aimed to address these limitations with inclusion of broader age ranges (recruiting adults from age 18), larger planned sample sizes, greater efforts towards inclusion of racially and ethnically diverse individuals, and the collection of information on gender identity and sexual orientation (Cook *et al.*, 2025a, Cook *et al.*, 2025b, The All of Us Research Program Investigators, 2019). Of particular relevance to this doctoral work, the recruitment of younger

adults in Our Future Health may permit the study of sex differences in cancer risk over longer exposure windows, allowing genetic susceptibility to be examined in relation to exposures established earlier in the life course rather than only those measured in mid-life. This is pertinent given evidence of rising incidence of early-onset cancer (diagnosed before age 50) for several cancer sites over recent decades, with global estimates suggesting a 79% increase between 1990 and 2019 (Zhao *et al.*, 2023). Its routine collection of gender identity alongside sex assigned at birth would also support analyses providing insights into sex differences in biology and disease (Cook *et al.*, 2025a, Cook *et al.*, 2025b). Although long follow-up will be required for incident cancer outcomes, combined with pooled multinational analyses, these features would strengthen the evidence base on the occupational, genetic, and hormonal contributions to the sex differences observed in this thesis and elsewhere.

## 7.5 Public Health Implications

The findings of this thesis have several implications for public health policy and cancer prevention. A central observation is that many of the sex differences in cancer incidence apparent in initial analyses attenuated substantially upon careful consideration of established risk factors, leaving a smaller, more focused set of cancers for which the sex differences remain unexplained and warrant further research. This finding is itself a public health insight and has practical implications for how prevention efforts are conceived, prioritised, and delivered, in particular in terms of highlighting the differences in risk profiles, that is, the distribution of established risk factors, between men and women. Moreover, there is evidence that interventions targeting the very risk factors that account for some of the major sex differences in cancer incidence, such as smoking and excess adiposity, are themselves subject to sex differences in effectiveness and delivery. Therefore, addressing sex differences in cancer incidence requires not only identifying the biological or behavioural factors that contribute to them, but also ensuring that prevention

strategies reach the groups in whom these risk factors are most prevalent or least effectively treated.

A clinical trial evaluating the effectiveness of physicians screening for and opportunistically intervening on obesity demonstrated that brief opportunistic interventions delivered by general practitioners in primary care were similarly effective in men and women, and that men and women who do enrol attend at similar rates and achieve comparable weight loss (Aveyard *et al.*, 2016). However, in routine UK primary care, general practitioners refer nine times as many women as men to community weight loss programmes (Aveyard *et al.*, 2016, Tudor *et al.*, 2021). When clinicians systematically endorsed and offered referrals in the trial setting, three-quarters of this sex-specific imbalance was removed (Tudor *et al.*, 2021). Given that excess adiposity contributes meaningfully to oesophageal adenocarcinoma, kidney cancer, and colorectal cancer, all of which had a male excess in initial analyses in this doctoral work, addressing this delivery gap could reduce sex differences.

A parallel set of issues applies to smoking. Although smoking prevalence is generally higher among men than women, women appear to have greater difficulty maintaining long-term abstinence following a quit attempt across both efficacy and effectiveness trials (Smith *et al.*, 2016). Mechanisms proposed include greater nicotine withdrawal severity across the menstrual cycle, higher behavioural (as distinct from pharmacological) dependence, and weight-control concerns acting as barriers to cessation (McKee *et al.*, 2016, Smith *et al.*, 2016). A meta-analysis of randomised trials found that varenicline (an FDA-approved treatment for nicotine dependence, alongside bupropion and nicotine replacement therapy) was approximately 46% more efficacious for women than men at end-of-treatment, and produced equal long-term abstinence rates across sexes, in contrast to other cessation pharmacotherapies which tend to favour men (McKee *et al.*, 2016). Smoking therefore remains a public health concern for both sexes. Among women it is most prevalent in high-income regions, where rates exceed twice the global average,

consistent with the higher burden of smoking-attributable cancers observed in these wealthier settings (Fink *et al.*, 2026). Context-specific public health strategies that account for these patterns may be particularly effective in reducing the cancer burden in women, given that global smoking cessation efforts have historically focused on men.

Beyond established risk factors, the monitoring of emerging exposures, with attention to their sex-specific distribution, is also important. The evolving landscape of nicotine delivery, including e-cigarettes and related products, may have consequences for cancer incidence that will only become apparent over decades, and may carry sex-specific patterns of uptake, biological effects, and cessation response. Surveillance of these exposures in younger adults, with attention to sex-specific patterns of use, may be valuable for informing legislation and cessation policy. This is particularly pertinent given the rising incidence of several cancers in younger people, most notably colorectal cancer, lung cancer in never-smokers, and certain gynaecological cancers. Therefore, capturing emerging exposures and over longer exposure windows than current middle-aged cohorts permit, will be crucial. Together, these considerations suggest that public health priorities should incorporate sex-specific distributions of exposures and develop context-specific strategies that reduce the cancer burden effectively in both sexes.

## 7.6 Conclusions

Overall, this thesis provides the first UK based comprehensive evaluation of sex differences in cancer incidence, demonstrating that while established risk factors explain some observed disparities, substantial differences persist for several cancer sites. Using data from a large UK cohort, four cancers were identified, oesophageal adenocarcinoma, bladder cancer, and rectal cancer (with higher incidence in men), and lung adenocarcinoma (with higher incidence in women), for which marked sex differences remained upon evaluation of contributions from risk factors and therefore merit further investigation. These findings highlight important gaps in

current understanding of cancer aetiology and underscore the importance of future research into sex-specific environmental and health-related exposures, as well as underlying biological and genetic mechanisms. Collectively, findings of this doctoral work may be a valuable reference for future investigations of sex differences in cancer aetiology.

**Table 7.1 Summary of hazard ratios (HR, 95% CI) comparing men to women for cancer sites with markedly[a] higher risk in men.**

| | Main cohort analyses | | | Sensitivity analyses (fully adjusted) | | | Adiposity substitution[d] (%Δ lnHR) | Interpretation and research implications |
|---|---|---|---|---|---|---|---|---|
| | Minimally adjusted[b] | Fully adjusted[b] | %Δ lnHR[c] | Never smokers | Alcohol <20g/day | BMI <25 kg/m$^2$ | | |
| Oesophageal adenocarcinoma [C15] | 6.0 (4.9-7.4) | 5.5 (4.2-7.1) | 5.5% | 5.1 (3.3-8.0) | 5.0 (3.6-7.1) | 4.7 (2.6-8.7) | WC: 4.4 (3.3-5.8); 13.5% WHR: 3.5 (2.6-4.8); 26.2% | Central adiposity traits partially contribute to excess risk in men, but there remains a large male predominance. Further work on mechanistic insights and molecular (hormonal and metabolic) risk factors is needed. |
| Bladder [C67] | 4.1 (3.5-4.7) | 3.5 (2.9-4.2) | 11.4% | 3.2 (2.3-4.4) | 3.6 (2.8-4.6) | 4.4 (3.0-6.4) | | Sex differences remain upon evaluation from risk factors. Further work on occupational exposures and genetic risk factors (Y chromosome) is needed. |
| Kidney [C64-C65] | 2.1 (1.9-2.4) | 1.8 (1.5-2.1) | 23.5% | 2.1 (1.6-2.7) | 1.7 (1.4-2.0) | 2.0 (1.4-2.8) | WC: 1.6 (1.3-1.9); 23.4% WHR: 1.4 (1.1-1.7); 48.5% | Central adiposity traits strongly contribute to the excess risk in men. Further work on molecular risk factors underpinning the waist-based measures is needed. |
| Rectum [C19-C20] | 2.1 (1.9-2.2) | 1.7 (1.5-2.0) | 27.5% | 1.9 (1.5-2.4) | 1.7 (1.4-2.0) | 1.5 (1.2-2.0) | Not associated with adiposity traits. | Sex differences remain upon evaluation from risk factors. Further work on risk factors specific to rectal cancer is needed. |

[a]Among the 15 cancers sites and 13 subtypes evaluated, these cancers have higher incidence in men compared to women after adjusting for a range of cancer risk factors.
[b]Minimally adjusted Cox PH models included stratifying for age at recruitment, deprivation, and region, and full models included additionally adjusting for ethnicity, educational qualifications, height, BMI, smoking, alcohol, and cancer site-specific risk factors.
[c]Percent change in ln(HR) from minimally to fully adjusted models. Calculated as $[\ln(HR_{\text{unadjusted}}) - \ln(HR_{\text{adjusted}})]/\ln(HR_{\text{unadjusted}}) \times 100]$.
[d]Analyses assessed sex differences in OAC and Kidney cancer risk by adiposity trait. Here, percent change in ln(HR) calculated relative to the no-adiposity trait adjustment model.

Bladder cancer was not followed up for adiposity analyses.

# Appendices

**Appendix Table 3A Exclusion criteria applied to UK Biobank analyses ..... 194**

**Appendix Table 3B ICD-10 codes for 15 cancers and 13 subtypes examined in pan-cancer analyses of sex differences in cancer incidence. ..... 195**

**Appendix Table 3C ICD-10 codes for cancer sites that make up malignant neoplasms of lip, oral cavity and pharynx. ..... 196**

**Appendix Table 3D: Histology codes for distinguishing cancer subtypes ..... 197**

**Appendix Table 4A: Prevalent disease definitions ..... 199**

**Appendix Table 4B: Cancer specific multivariable-adjusted Cox proportional hazards models comparing risk of cancer incidence in men versus women in the highest qualification attainment stratum (Prof Q/NVQ/HND/HNC/Degree or other professional qualifications) in the UK Biobank. ..... 201**

**Appendix Table 4C: Cox proportional hazards models comparing risk of cancer incidence in men versus women with additional adjustment for non-linear height (height$^2$) added to the model. ..... 203**

**Appendix Table 4D: Cox proportional hazards model comparing risk of cancer incidence in men versus women with additional adjustment for physical activity (METS/week). ..... 204**

**Appendix Table 4E Cancer site-specific Cox models rerun with stratification by 3 broader geographic regions (England, Wales and Scotland) ..... 205**

**Appendix Table 4F Percentage change in the natural logarithm of hazard ratios (%Δln(HR)) from minimally adjusted to fully adjusted Cox PH models. ..... 206**

**Appendix Table 4G HRs and 95% CIs comparing risk of cancer in men to women (Table 4.1 with unformatted p-values) ..... 207**

**Appendix Table 5A Hazard ratios (HRs) and 95% confidence intervals (CIs) comparing rectum cancer risk [C19-C20] risk by adiposity measures, using data from 249,602 women and 213,163 men in the UK Biobank. .....212**

**Appendix Table 5B Hazard ratios (95% confidence intervals) comparing men with women for risk of oesophageal adenocarcinoma (ICD-10 C15), adjusted for one adiposity trait at a time. .....214**

**Appendix Table 5C Hazard ratios (95% confidence intervals) comparing men with women for risk of kidney cancer [ICD-10 C64-C65], adjusted for one adiposity trait at a time.....215**

**Appendix Table 7A: Summary of hazard ratios (HR, 95% CI) comparing men to women for all[a] cancer sites with higher risk in men. .....216**

**Appendix Table 7B: Summary of hazard ratios (HR, 95% CI) comparing men to women, for all cancer sites with higher risk in women.....218**

**Appendix Table 3A Exclusion criteria applied to UK Biobank analyses**

| Stage | Exclusion criterion | Number excluded | Participants remaining |
|---|---|---|---|
| Initial cohort | UK Biobank participants recruited at baseline | – | **502,386** |
| Core exclusions | Missing key baseline covariates (sex, region, age, height, BMI) | 3,105 | 499,281 |
| | Prevalent cancer at baseline (excluding non-melanoma skin cancer, ICD-10 C44) | 28,143 | 471,138 |
| | Discordance between self-reported and genetically inferred sex | 367 | 470,771 |
| **Chapter 4: Pan-cancer Analyses of Sex Differences in Cancer Incidence** | | | |
| Primary analytical cohort | Eligible participants for main analyses | – | **470,771** |
| Cancer-specific analyses | History of partial or complete gastrectomy | 51 | **470,720*** |
| **Chapter 4 Sensitivity Checks: Pan-cancer Analyses Restricted to Select Cohorts** | | | |
| Never smokers | Current and former smokers | 213,330 | 257,390 |
| Light alcohol drinkers | Alcohol intake >20g/day | 169,211 | 301,509 |
| Normal BMI | BMI>25kg/m$^2$ | 317,874 | 152,846 |
| Highest educational qualification | Participants without the highest educational qualification | 191,241 | 279,479 |
| **Chapter 5: Investigating Adiposity Traits in Relation to Sex Differences** | | | |
| Analytical cohort | Missing body composition measures (waist or hip circumference, body fat or trunk fat measures) | 7,955 | **462,765** |
| **Chapter 6: Lung adenocarcinoma in Never Smokers** | | | |
| Analytical cohort | Current or former smokers | 213,330 | **257,390** |

*The final cohort size for the cancer-specific multivariable adjusted models included 470,720 participants, and subsequent exclusions for Chapter 4 sensitivity analyses, Chapter 5, and Chapter 6 analyses were applied to this cohort.

**Appendix Table 3B ICD-10 codes for 15 cancers and 13 subtypes examined in pan-cancer analyses of sex differences in cancer incidence.**

| ICD-10 codes | Cancer site |
|---|---|
| C00-C14 | Lip, oral cavity and pharynx[a] |
| C15 | Oesophagus |
| C15 | Oesophageal adenocarcinoma |
| C15 | Oesophageal squamous cell carcinoma |
| C16 | Stomach |
| C16.0 | Gastric cardia |
| C16.1-C16.6 | Gastric non-cardia |
| C18-C21 | Colorectum |
| C18 | Colon |
| C19-C20 | Rectum |
| C21 | Anus |
| C22 | Liver |
| C23 | Gallbladder |
| C25 | Pancreas |
| C34 | Lung |
| C34 | Lung adenocarcinoma |
| C34 | Lung small cell carcinoma |
| C34 | Lung squamous cell carcinoma |
| C43 | Melanoma |
| C50 | Breast |
| C64-C65 | Kidney |
| C67 | Bladder |
| C73 | Thyroid |
| C81-C96 | Lymphatic and haematopoietic cancers |
| C81 | Hodgkin lymphoma |
| C82-85 | Non-Hodgkin lymphoma |
| C88, C90 | Multiple myeloma |
| C91-95 | Leukaemia |

[a]Referred to as cancer of the 'oral cavity' throughout the thesis (see Appendix Table 3C)

**Appendix Table 3C ICD-10 codes for cancer sites that make up malignant neoplasms of lip, oral cavity and pharynx.**

| ICD-10 codes (C00-C14) | Malignant neoplasms of the |
|---|---|
| C00 | Lip |
| C01 | Base of tongue |
| C02 | Other and unspecified parts of tongue |
| C03 | Gum |
| C04 | Floor of mouth |
| C05 | Palate |
| C06 | Other and unspecific parts of mouth |
| C07 | Parotid gland |
| C08 | Other and unspecified major salivary glands |
| C09 | Tonsil |
| C10 | Oropharynx |
| C11 | Nasopharynx |
| C12 | Piriform sinus |
| C13 | Hypopharynx |
| C14 | Other and ill-defined sites in the lip, oral cavity and pharynx |

**Appendix Table 3D: Histology codes for distinguishing cancer subtypes**

| **WHO ICD-O-3 Oesophageal cancer [ICD C15]** | |
|---|---|
| **Oesophageal adenocarcinoma** | 8140/3 Adenocarcinoma NOS<br>8144/3 Adenocarcinoma, intestinal type<br>8145/3 Adenocarcinoma, diffuse type<br>8246/3 Neuroendocrine carcinoma, NOS<br>8260/3 Papillary, NOS<br>8310/3 Clear cell adenocarcinoma, NOS<br>8312/3 Adenocarcinoma, renal cell type,<br>8323/3 Mixed cell adenocarcinoma<br>8402/3 Nodular hidradenoma, malignant<br>8440/3 Cystadenocarcinoma, NOS<br>8461/3 Serous surface papillary carcinoma<br>8480/3 Mucinous adenocarcinoma<br>8481/3 Mucin-producing adenocarcinoma<br>8490/3 Signet ring cell adenocarcinoma<br>8500/3 Infiltrating duct adenocarcinoma<br>8543/3 Intraductal carcinoma<br>8560/3 Adenocarcinoma and epidermoid carcinoma, mixed |
| **Oesophageal squamous cell carcinoma** | 8070/3 Epidermoid NOS<br>8071/3 Squamous cell carcinoma keratinizing, NOS<br>8081/3 Intraepidermal squamous cell carcinoma |
| **WHO ICD-O-3 Lung cancer [ICD C34] histological subtypes**<br>**These histology codes had no lung cancer cases in the UK Biobank cohort.* | |
| **Lung adenocarcinoma** | 8140/3 Adenocarcinoma, NOS<br>8140/6 NOS, metastatic<br>*8190/3 trabecular adenocarcinoma*<br>*8211/3 tubular adenocarcinoma*<br>8244/3 carcinoid and adenocarcinoma<br>8250/3 bronchiolar (C34.1), bronchiolo-alveolar, NOS<br>8251/3 alveolar adenocarcinoma<br>*8252/3 Bronchiolo-alveolar carcinoma, non-mucinous*<br>*8253/3 Bronchio-alveolar goblet cell type adenocarcinoma*<br>*8254/3 Bronchio-alveolar adenocarcinoma: clara cell and goblet cell type, mixed mucinous and non-mucinous, type II pneumocyte and goblet cell type*<br>8255/3 adenocarcinoma combined with other types of carcinomas<br>8260/3 Papillary adenocarcinoma, NOS<br>*8290/3 Oxyphilic adenocarcinoma*<br>8310/3 clear cell adenocarcinoma, NOS<br>8323/3 Mixed cell adenocarcinoma<br>8440/3 Cystadenocarcinoma, NOS<br>8480/3 mucinous adenocarcinoma, NOS<br>8481/3 mucin-producing, mucin secreting<br>8490/3 Signet ring cell adenocarcinoma<br>8550/3 acinar adenocarcinoma<br>*8570/3 adenocarcinoma with squamous metaplasia* |

| | |
|---|---|
| | *8571/3 adenocarcinoma with cartilaginous and osseous metaplasia*<br>*8572/3 adenocarcinoma with spindle cell metaplasia*<br>*8573/3 adenocarcinoma with apocrine metaplasia*<br>8574/3 adenocarcinoma with neuroendocrine differentiation<br>*8576/3 Hepatoid adenocarcinoma* |
| **Lung small cell carcinoma** | 8041/3 small cell carcinoma, NOS<br>8042/3 oat cell carcinoma<br>8045/3 combined small cell carcinoma (combined small cell-adenocarcinoma, combined small cell-large cell carcinoma, combined small cell-squamous cell carcinoma |
| **Lung squamous cell carcinoma** | *8051/3 verrucous squamous cell carcinoma*<br>*8052/3 papillary squamous cell carcinoma*<br>8070/3 squamous cell carcinoma, NOS<br>8071/3 squamous cell carcinoma, keratinizing, NOS<br>8072/3 squamous cell carcinoma, large cell, nonkeratinizing, NOS<br>*8073/3 squamous cell carcinoma, small cell, nonkeratinizing*<br>8074/3 squamous cell carcinoma, spindle cell<br>*8075/3 squamous cell carcinoma, adenoid*<br>*8076/3 squamous cell carcinoma, microinvasive*<br>8083/3 basaloid squamous cell carcinoma<br>8084/3 squamous cell carcinoma, clear cell type |

**References used for the classification along with the ICD-O-3 codes**

World Health Organization (2013) International Classification of diseases for Oncology, 3rd Edition: First Revision. https://apps.who.int/iris/bitstream/handle/10665/96612/9789241548496_eng.pdf.

Ruhl JL, Callaghan C, Schussler N (eds.) Summary Stage 2018: Codes and Coding Instructions, National Cancer Institute, Bethesda, MD, 2022.

Watts EL, Perez-Cornago A, Knuppel A, Tsilidis KK, Key TJ, Travis RC, Prospective analyses of testosterone and sex hormone-binding globulin with the risk of 19 types of cancer in men and postmenopausal women in the UK Biobank. *Int. J. Cancer*. 2021; 149: 573–584. https://doi.org/10.1002/ijc.33555

Pirie, K., Peto, R., Green, J., Reeves, G.K., Beral, V. (2016), Lung cancer in never smokers in the UK Million Women Study. Int. J. Cancer, 139: 347-354. https://doi.org/10.1002/ijc.30084

Jason Y Y Wong, ScD, Bryan A Bassig, PhD, Erikka Loftfield, PhD, Wei Hu, PhD, Neal D Freedman, PhD, Bu-tian Ji, PhD, Paul Elliott, PhD, Debra T Silverman, ScD, Stephen J Chanock, MD, Nathaniel Rothman, MD, Qing Lan, MD, PhD, White Blood Cell Count and Risk of Incident Lung Cancer in the UK Biobank, *JNCI Cancer Spectrum*, Volume 4, Issue 2, April 2020, pkz102, https://doi.org/10.1093/jncics/pkz102

**Appendix Table 4A: Prevalent disease definitions**

| Prevalent disease/ admissions cause | Outcome definition using ICD-10[a] | Outcome definition using ICD-9 | Relevant procedure code using OPSC-4 definitions[a] | Touchscreen criteria[b] | Medication criteria[c] | Interview codes for prevalent diseases[d] | Total cases [in women] |
|---|---|---|---|---|---|---|---|
| Alcoholic Liver Disease | K70 | | | | | 1604 | 498 [103] |
| Diabetes Mellitus | E10, E11, E12, E13, E14 | 250 | | 2443 | 6177 | 1220 1222 1223 | 25,156 [9,760] |
| Gallbladder disease | K80, K81, K82.9 | | | | | 1161 | 9,773 [7,327] |
| Gastrectomy (partial or complete)[e] | | | G271 G272 G273 G274 G275 G276 G277 G278 G279 G281 G282 G283 G284 G285 G286 G287 G288 G289 | | | | 51 |
| Gastro-oesophagus reflux disease[f] | K21 | 530.11<br>530.81 | | | | | 13,925 [7,037] |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Goiter | E049 | | | | | | 584 [452] |
| Human Immunodeficiency Virus | B20 B21 B22 B23 B24 | 042 | | | | 1439 | 455 [<100] |
| Hypertension | I10 | | | | | 1065 | 126,410 [59,481] |
| Inflammatory Bowel Disease | K50 K51 K52.3 | 555 556 | | | | 1462 1663 1461 | 4,580 [2,668] |
| Non-alcoholic fatty liver disease | K75.8 K76 K74.6 | | | | | | 1,174 [561] |

[a] Based on OPSC-4 definition and UKB variable 20004.

[b] Used for identifying prevalent diabetes (data-field 2443); participants were asked at recruitment 'Has a doctor ever told you that you have diabetes? (yes/no/do not know/prefer not to answer). Participants that selected 'yes' were classed as those having diabetes.

[c] Used for identifying prevalent diabetes (data field 6177); participants were asked at recruitment 'Do you regularly take any of the following medications? (you can select more than one answer) (cholesterol lowering medication/blood pressure medication/insulin/none of the above/do not know). Participants that selected 'insulin' were classed as those having diabetes.

[d] Based on ICD-10 definition and UKB variable 20002. See text under prevalent diseases.

[e] Participants with complete or partial gastrectomy were excluded for the cancer-specific multivariable adjusted analyses.

[f] As part of OAC adjustment, prevalent Barrett's oesophagus was also considered for assessment using ICD-10 code K22.7 (primary and secondary diagnoses recorded prior to recruitment) and self-reported non-cancer illness code 1139, however, no prevalent cases were identified at baseline.

**Appendix Table 4B: Cancer specific multivariable-adjusted Cox proportional hazards models comparing risk of cancer incidence in men versus women in the highest qualification attainment stratum (Prof Q/NVQ/HND/HNC/Degree or other professional qualifications) in the UK Biobank.**

| | Women (%) | Men (%) | Total | HR (95% CI) | P value | Main analysis results |
|---|---|---|---|---|---|---|
| Oral cavity [C00-C14] | 178 (34.5) | 338 (65.5) | 516 | **1.93 (1.48-2.52)** | 1.28e-06 | 2.06 (1.69-2.51) |
| Oesophagus [C15] | 126 (24.5) | 388 (75.5) | 514 | **2.92 (2.21-3.87)** | 7.17e-14 | 2.89 (2.36-3.55) |
| Oesophageal adenocarcinoma [C15] | 53 (14.6) | 310 (85.4) | 363 | **6.27 (4.32-9.09)** | 3.94e-22 | 5.45 (4.18-7.12) |
| Oesophageal SCC [C15] | 65 (52) | 60 (48) | 125 | **0.6 (0.35-1.02)** | 0.061 | 0.81 (0.55-1.2) |
| Stomach [C16] | 89 (30.6) | 202 (69.4) | 291 | **1.94 (1.35-2.79)** | 0.000328 | 2.34 (1.83-2.97) |
| Gastric cardia [C16.0] | 26 (18.6) | 114 (81.4) | 140 | **4.04 (2.28-7.16)** | 1.63e-06 | 3.65 (2.48-5.38) |
| Gastric non-cardia [C16.1-C16.6] | 35 (39.8) | 53 (60.2) | 88 | *[insufficient data]* | *[insufficient data]* | 1.69 (1.09-2.6) |
| Colorectum [C18-C20] | 1217 (40.9) | 1761 (59.1) | 2978 | **1.14 (1.02-1.28)** | 0.02 | 1.22 (1.12-1.33) |
| Colon [C18] | 853 (43.7) | 1099 (56.3) | 1952 | **1 (0.87-1.14)** | 0.973 | 1.04 (0.94-1.15) |
| Rectum [C19-C20] | 371 (35.7) | 669 (64.3) | 1040 | **1.5 (1.24-1.81)** | 3.51e-05 | 1.7 (1.47-1.96) |
| Anus [C21] | 66 (61.1) | 42 (38.9) | 108 | **0.44 (0.24-0.79)** | 0.006 | 0.41 (0.26-0.64) |
| Liver [C22] | 88 (30.7) | 199 (69.3) | 287 | **1.83 (1.26-2.64)** | 0.001 | 1.91 (1.48-2.47) |
| Gallbladder [C23] | 31 (60.8) | 20 (39.2) | 51 | *[insufficient data]* | *[insufficient data]* | 0.55 (0.31-0.97) |
| Pancreas [C25] | 284 (44.3) | 357 (55.7) | 641 | **0.95 (0.75-1.2)** | 0.656 | 1.04 (0.87-1.25) |
| Lung [C34] | 709 (47) | 800 (53) | 1509 | **0.92 (0.79-1.07)** | 0.284 | 0.93 (0.84-1.03) |
| Lung adenocarcinoma [C34] | 355 (52.5) | 321 (47.5) | 676 | **0.69 (0.55-0.87)** | 0.002 | 0.72 (0.62-0.84) |
| Lung SCC [C34] | 96 (33.6) | 190 (66.4) | 286 | **1.57 (1.1-2.26)** | 0.014 | 1.41 (1.14-1.75) |

**Appendix Table 4B: Cancer specific multivariable-adjusted Cox proportional hazards models comparing risk of cancer incidence in men versus women in the highest qualification attainment stratum (Prof Q/NVQ/HND/HNC/Degree or other professional qualifications) in the UK Biobank.**

| | Women (%) | Men (%) | Total | HR (95% CI) | P value | Main analysis results |
|---|---|---|---|---|---|---|
| Lung small cell carcinoma [C34] | 58 (43.9) | 74 (56.1) | 132 | **1.21 (0.71-2.04)** | 0.485 | 1.19 (0.86-1.64) |
| Melanoma [C43] | 763 (47.5) | 844 (52.5) | 1607 | **0.87 (0.75-1.01)** | 0.076 | 0.97 (0.86-1.09) |
| Breast [C50] | 4892 (99.3) | 36 (0.7) | 4928 | **0.006 (0.004-0.008)** | 6.7e-203 | 0.005 (0.004-0.007) |
| Kidney [C64-65] | 238 (32.9) | 485 (67.1) | 723 | **1.82 (1.45-2.29)** | 2.11e-07 | 1.77 (1.51-2.09) |
| Bladder [C67] | 97 (17.5) | 458 (82.5) | 555 | **3.93 (2.95-5.25)** | 1.47e-20 | 3.47 (2.85-4.24) |
| Thyroid [C73] | 156 (70.3) | 66 (29.7) | 222 | **0.36 (0.24-0.55)** | 1.92e-06 | 0.36 (0.26-0.49) |
| LHTC [C81-C96] | 1019 (41.4) | 1443 (58.6) | 2462 | **1.18 (1.04-1.33)** | 0.008 | 1.22 (1.11-1.33) |
| Hodgkin lymphoma [C81] | 31 (45.6) | 37 (54.4) | 68 | *[insufficient data]* | *[insufficient data]* | 1.29 (0.77-2.17) |
| Non-Hodgkin lymphoma [C82-85] | 504 (43.3) | 660 (56.7) | 1164 | **1.07 (0.9-1.27)** | 0.467 | 1.04 (0.91-1.19) |
| Multiple myeloma [C90 & C88] | 226 (42.3) | 308 (57.7) | 534 | **1.23 (0.95-1.6)** | 0.113 | 1.35 (1.10-1.65) |
| Leukaemia [C91-C95] | 251 (37) | 427 (63) | 678 | **1.33 (1.06-1.68)** | 0.015 | 1.43 (1.21-1.69) |

This sensitivity analysis was performed on a multivariable-adjusted model that included cancer-specific, and baseline set of covariates. Covariate selection has been described in text and supplementary tables. Cancer endpoints were excluded from the analysis if total counts were less than 100. **279,479** participants (51.3% women) were categorized in the highest qualification stratum based on their self-reports and it included individuals with 'College or university degree', 'NVQ or HND or HNC or equivalent', 'Other professional qualifications e.g.: nursing, teaching'. Within this strata, 17,392 participants had malignant cancers, with 9,953 cases (57.2%) in women.

The main analyses results for all participants (HRs and 95% CIs from multivariable-adjusted models with cancer specific covariates) are also shown for comparison.

LHTC lymphatic and haematopoietic cancers; SCC squamous cell carcinoma.

**Appendix Table 4C: Cox proportional hazards models comparing risk of cancer incidence in men versus women with additional adjustment for non-linear height (height$^2$) added to the model.**

| | HR (95% CI) | P value | Main analysis results |
|---|---|---|---|
| Oral cavity [C00-C14] | **2.02 (1.66-2.47)** | 4.07e-12 | 2.06 (1.69-2.51) |
| Oesophagus [C15] | **2.74 (2.22-3.38)** | 1.03e-20 | 2.89 (2.36-3.55) |
| Oesophageal adenocarcinoma [C15] | **5.02 (3.78-6.67)** | 8.98e-29 | 5.45 (4.18-7.12) |
| Oesophageal SCC [C15] | **0.81 (0.55-1.19)** | 0.291 | 0.81 (0.55-1.2) |
| Stomach [C16] | **2.22 (1.74-2.84)** | 1.82e-10 | 2.34 (1.83-2.97) |
| Gastric cardia [C16.0] | **3.73 (2.48-5.62)** | 2.87e-10 | 3.65 (2.48-5.38) |
| Gastric non-cardia [C16.1-C16.6] | **1.58 (1.03-2.42)** | 0.036 | 1.69 (1.09-2.6) |
| Colorectum [C18-C20] | **1.22 (1.12-1.32)** | 4.17e-06 | 1.22 (1.12-1.33) |
| Colon [C18] | **1.04 (0.94-1.15)** | 0.484 | 1.04 (0.94-1.15) |
| Rectum [C19-C20] | **1.69 (1.46-1.96)** | 1.92e-12 | 1.7 (1.47-1.96) |
| Anus [C21] | **0.42 (0.27-0.66)** | 1.44e-04 | 0.41 (0.26-0.64) |
| Liver [C22] | **1.89 (1.46-2.46)** | 1.78e-06 | 1.91 (1.48-2.47) |
| Gallbladder [C23] | **0.56 (0.32-1.00)** | 0.049 | 0.55 (0.31-0.97) |
| Pancreas [C25] | **1.05 (0.88-1.25)** | 0.606 | 1.04 (0.87-1.25) |
| Lung [C34] | **0.93 (0.84-1.02)** | 0.134 | 0.93 (0.84-1.03) |
| Lung adenocarcinoma [C34] | **0.72 (0.62-0.84)** | 2.49e-05 | 0.72 (0.62-0.84) |
| Lung SCC [C34] | **1.41 (1.14-1.76)** | 0.002 | 1.41 (1.14-1.75) |
| Lung small cell carcinoma [C34] | **1.18 (0.85-1.62)** | 0.316 | 1.19 (0.86-1.64) |
| Melanoma [C43] | **0.97 (0.86-1.09)** | 0.567 | 0.97 (0.86-1.09) |
| Breast [C50] | **0.005 (0.004-0.007)** | 6.30e-302 | 0.005 (0.004-0.007) |
| Kidney [C64-65] | **1.8 (1.52-2.13)** | 8.23e-12 | 1.77 (1.51-2.09) |
| Bladder [C67] | **3.78 (3.05-4.67)** | 1.33e-34 | 3.47 (2.85-4.24) |
| Thyroid [C73] | **0.36 (0.26-0.5)** | 1.28e-09 | 0.36 (0.26-0.49) |
| LHTC [C81-C96] | **1.21 (1.11-1.33)** | 3.60e-05 | 1.22 (1.11-1.33) |
| Hodgkin lymphoma [C81] | **1.24 (0.75-2.03)** | 0.397 | 1.29 (0.77-2.17) |
| Non-Hodgkin lymphoma [C82-85] | **1.03 (0.9-1.18)** | 0.635 | 1.04 (0.91-1.19) |
| Multiple myeloma [C90 & C88] | **1.37 (1.12-1.68)** | 0.003 | 1.35 (1.10-1.65) |
| Leukaemia [C91-C95] | **1.43 (1.2-1.69)** | 5.02e-05 | 1.43 (1.21-1.69) |

Cox PH models are adjusted for ethnicity, qualifications, height, height$^2$, BMI, smoking, alcohol consumption, and cancer specific risk factors, and stratified for age group, Townsend deprivation index, and region. The main analyses results for all participants (HRs and 95% CIs from multivariable-adjusted models with cancer specific covariates) are also shown for comparison.

LHTC lymphatic and haematopoietic cancers; SCC squamous cell carcinoma.

**Appendix Table 4D: Cox proportional hazards model comparing risk of cancer incidence in men versus women with additional adjustment for physical activity (METS/week).**

| | HR (95% CI) | P value | Main analysis results |
|---|---|---|---|
| Oral cavity [C00-C14] | **2.12 (1.74-2.57)** | 5.23e-14 | 2.06 (1.69-2.51) |
| Oesophagus [C15] | **2.91 (2.37-3.57)** | 1.34e-24 | 2.89 (2.36-3.55) |
| Oesophageal adenocarcinoma [C15] | **5.47 (4.19-7.14)** | 8.16e-36 | 5.45 (4.18-7.12) |
| Oesophageal SCC [C15] | **0.86 (0.59-1.27)** | 0.458 | 0.81 (0.55-1.2) |
| Stomach [C16] | **2.3 (1.81-2.93)** | 1.39e-11 | 2.34 (1.83-2.97) |
| Gastric cardia [C16.0] | **3.62 (2.46-5.33)** | 6.52e-11 | 3.65 (2.48-5.38) |
| Gastric non-cardia [C16.1-C16.6] | **1.69 (1.09-2.6)** | 0.018 | 1.69 (1.09-2.6) |
| Colorectum [C18-C20] | **1.22 (1.13-1.33)** | 1.91e-06 | 1.22 (1.12-1.33) |
| Colon [C18] | **1.04 (0.94-1.15)** | 0.43 | 1.04 (0.94-1.15) |
| Rectum [C19/C20] | **1.7 (1.47-1.97)** | 4.40e-13 | 1.7 (1.47-1.96) |
| Anus [C21] | **0.41 (0.26-0.64)** | 8.55e-05 | 0.41 (0.26-0.64) |
| Liver [C22] | **1.94 (1.5-2.5)** | 4.50e-07 | 1.91 (1.48-2.47) |
| Gallbladder [C23] | **0.57 (0.32-1.02)** | 0.058 | 0.55 (0.31-0.97) |
| Pancreas [C25] | **1.03 (0.87-1.24)** | 0.707 | 1.04 (0.87-1.25) |
| Lung [C34] | **0.94 (0.85-1.04)** | 0.243 | 0.93 (0.84-1.03) |
| Lung adenocarcinoma [C34] | **0.72 (0.62-0.84)** | 2.94e-05 | 0.72 (0.62-0.84) |
| Lung squamous cell carcinoma [C34] | **1.45 (1.17-1.8)** | 6.50e-04 | 1.41 (1.14-1.75) |
| Lung small cell carcinoma [C34] | **1.21 (0.88-1.67)** | 0.244 | 1.19 (0.86-1.64) |
| Melanoma [C43] | **0.96 (0.86-1.09)** | 0.541 | 0.97 (0.86-1.09) |
| Breast [C50] | **0.005 (0.004-0.007)** | 0.00e+00 | 0.005 (0.004-0.007) |
| Kidney [C64-65] | **1.78 (1.51-2.1)** | 7.33e-12 | 1.77 (1.51-2.09) |
| Bladder [C67] | **3.46 (2.84-4.22)** | 1.67e-34 | 3.47 (2.85-4.24) |
| Thyroid [C73] | **0.36 (0.26-0.49)** | 4.76e-10 | 0.36 (0.26-0.49) |
| LHTC [C81-C96] | **1.22 (1.11-1.34)** | 2.16e-05 | 1.22 (1.11-1.33) |
| Hodgkin lymphoma [C81] | **1.29 (0.77-2.17)** | 0.328 | 1.29 (0.77-2.17) |
| Non-Hodgkin lymphoma [C82-85] | **1.04 (0.91-1.19)** | 0.569 | 1.04 (0.91-1.19) |
| Multiple myeloma [C90 & C88] | **1.36 (1.11-1.66)** | 0.003 | 1.35 (1.10-1.65) |
| Leukaemia [C91-C95] | **1.43 (1.21-1.69)** | 3.85e-05 | 1.43 (1.21-1.69) |

Physical activity was self-reported in the UK Biobank; participants answered questions about their physical activity levels at recruitment. Based on IPAQ guidelines, we categorized participants' physical activity levels (based on their metabolic equivalent (METS) per week) as low, moderate, high, and missing/unknown. Cox PH models are adjusted for ethnicity, qualifications, height, BMI, smoking, alcohol consumption, physical activity levels, and cancer specific risk factors, and stratified by age group, Townsend deprivation index, and region. The main analysis results (HRs and 95% CIs from multivariable-adjusted models with cancer specific covariates) are shown for comparison. LHTC lymphatic and haematopoietic cancers; SCC squamous cell carcinoma.

**Appendix Table 4E Cancer site-specific Cox models rerun with stratification by 3 broader geographic regions (England, Wales and Scotland)**

| | Fully adjusted HRs (95% CI) | P value | Main analysis results |
|---|---|---|---|
| Oral cavity [C00-C14] | **2.07 (1.69-2.51)** | 3.45e-13 | 2.06 (1.69-2.51) |
| Oesophagus [C15] | **2.93 (2.39-3.60)** | 4.41e-25 | 2.89 (2.36-3.55) |
| OAC [C15] | **5.57 (4.27-7.26)** | 7.08e-37 | 5.45 (4.18-7.12) |
| Oesophageal SCC [C15] | **0.81 (0.55-1.19)** | 0.291 | 0.81 (0.55-1.2) |
| Stomach [C16] | **2.36 (1.86-3.00)** | 2.42e-12 | 2.34 (1.83-2.97) |
| Gastric cardia [C16.0] | **3.74 (2.54-5.49)** | 1.78e-11 | 3.65 (2.48-5.38) |
| Gastric non-cardia [C16.1-C16.6] | **1.75 (1.14-2.69)** | 0.010 | 1.69 (1.09-2.6) |
| Colorectum [C18-C20] | **1.21 (1.12-1.32)** | 2.78e-06 | 1.22 (1.12-1.33) |
| Colon [C18] | **1.04 (0.94-1.15)** | 0.469 | 1.04 (0.94-1.15) |
| Rectum [C19-C20] | **1.69 (1.46-1.96)** | 6.09e-13 | 1.7 (1.47-1.96) |
| Anus [C21] | **0.43 (0.27-0.66)** | 1.56e-4 | 0.41 (0.26-0.64) |
| Liver [C22] | **1.93 (1.49-2.49)** | 5.28e-07 | 1.91 (1.48-2.47) |
| Gallbladder [C23] | **0.59 (0.33-1.04)** | 0.071 | 0.55 (0.31-0.97) |
| Pancreas [C25] | **1.04 (0.87-1.24)** | 0.688 | 1.04 (0.87-1.25) |
| Lung [C34] | **0.94 (0.84-1.02)** | 0.199 | 0.93 (0.84-1.03) |
| Lung adenocarcinoma [C34] | **0.72 (0.62-0.84)** | 2.28e-05 | 0.72 (0.62-0.84) |
| Lung SCC [C34] | **1.45 (1.17-1.79)** | 0.0006 | 1.41 (1.14-1.75) |
| Lung small cell carcinoma [C34] | **1.21 (0.88-1.66)** | 0.242 | 1.19 (0.86-1.64) |
| Melanoma [C43] | **0.96 (0.86-1.08)** | 0.537 | 0.97 (0.86-1.09) |
| Breast [C50] | **0.005 (0.004-0.007)** | *p<1e-310** | 0.005 (0.004-0.007) |
| Kidney [C64-65] | **1.79 (1.52-2.11)** | 2.69e-12 | 1.77 (1.51-2.09) |
| Bladder [C67] | **3.44 (2.82-4.19)** | 2.45e-34 | 3.47 (2.85-4.24) |
| Thyroid [C73] | **0.36 (0.26-0.49)** | 5e-10 | 0.36 (0.26-0.49) |
| LHTC [C81-C96] | **1.22 (1.11-1.33)** | 2.47e-05 | 1.22 (1.11-1.33) |
| Hodgkin lymphoma [C81] | **1.27 (0.76-2.11)** | 0.361 | 1.29 (0.77-2.17) |
| Non-Hodgkin lymphoma [C82-85] | **1.04 (0.90-1.18)** | 0.562 | 1.04 (0.91-1.19) |
| Multiple myeloma [C90 & C88] | **1.35 (1.10-1.65)** | 0.003 | 1.35 (1.10-1.65) |
| Leukaemia [C91-C95] | **1.43 (1.20-1.69)** | 3.58e-05 | 1.43 (1.21-1.69) |

Fully adjusted Cox models were rerun with stratification by three broader geographic regions (England, Wales, and Scotland), rather than the ten-region classification used in the primary analyses (see Chapter 3, section 3.9.2)

*p-value was not retrievable further.

OAC oesophageal adenocarcinoma; LHTC lymphatic and haematopoietic cancers; SCC squamous cell carcinoma,

**Appendix Table 4F Percentage change in the natural logarithm of hazard ratios (%Δln(HR)) from minimally adjusted to fully adjusted Cox PH models.**

| | Minimally adjusted | Fully adjusted | %Δln(HR)* |
|---|---|---|---|
| Oral cavity [C00-C14] | 2.34 (2.04-2.67) | 2.06 (1.69-2.51) | 15.0 |
| Oesophagus [C15] | 3.3 (2.86-3.82) | 2.89 (2.36-3.55) | 11.1 |
| OAC [C15] | 6.01 (4.9-7.37) | 5.45 (4.18-7.12) | 5.5 |
| Oesophageal SCC [C15] | 1.07 (0.83-1.37) | 0.81 (0.55-1.2) | *NA* |
| Stomach [C16] | 2.55 (2.16-3.01) | 2.34 (1.83-2.97) | 9.2 |
| Gastric cardia [C16.0] | 4.17 (3.15-5.53) | 3.65 (2.48-5.38) | 9.3 |
| Gastric non-cardia [C16.1-C16.6] | 1.74 (1.31-2.33) | 1.69 (1.09-2.6) | 5.3 |
| Colorectum [C18-C20] | 1.55 (1.47-1.64) | 1.22 (1.12-1.33) | 54.6 |
| Colon [C18] | 1.34 (1.25-1.43) | 1.04 (0.94-1.15) | 86.6 |
| Rectum [C19-C20] | 2.08 (1.89-2.29) | 1.7 (1.47-1.96) | 27.5 |
| Anus [C21] | 0.62 (0.46-0.83) | 0.41 (0.26-0.64) | -86.5 |
| Liver [C22] | 2.34 (1.97-2.79) | 1.91 (1.48-2.47) | 23.9 |
| Gallbladder [C23] | 0.64 (0.44-0.94) | 0.55 (0.31-0.97) | -34.0 |
| Pancreas [C25] | 1.38 (1.23-1.55) | 1.04 (0.87-1.25) | 87.8 |
| Lung [C34] | 1.23 (1.16-1.32) | 0.93 (0.84-1.03) | 135.1 |
| Lung adenocarcinoma [C34] | 0.98 (0.89-1.09) | 0.72 (0.62-0.84) | *NA* |
| Lung SCC [C34] | 1.94 (1.67-2.24) | 1.41 (1.14-1.75) | 48.2 |
| Lung small cell carcinoma [C34] | 1.42 (1.15-1.75) | 1.19 (0.86-1.64) | 50.4 |
| Melanoma [C43] | 1.22 (1.13-1.31) | 0.97 (0.86-1.09) | 115.3 |
| Breast [C50] | 0.007 (0.006-0.009) | 0.005 (0.004-0.007) | -6.8 |
| Kidney [C64-65] | 2.11 (1.89-2.36) | 1.77 (1.51-2.09) | 23.5 |
| Bladder [C67] | 4.07 (3.53-4.71) | 3.47 (2.85-4.24) | 11.4 |
| Thyroid [C73] | 0.43 (0.34-0.53) | 0.36 (0.26-0.49) | -21.1 |
| LHTC [C81-C96] | 1.51 (1.42-1.60) | 1.22 (1.11-1.33) | 51.7 |
| Hodgkin lymphoma [C81] | 1.31 (0.93-1.84) | 1.29 (0.77-2.17) | 5.7 |
| Non-Hodgkin lymphoma [C82-85] | 1.35 (1.24-1.47) | 1.04 (0.91-1.19) | 86.9 |
| Multiple myeloma [C90 & C88] | 1.54 (1.35-1.76) | 1.35 (1.10-1.65) | 30.5 |
| Leukaemia [C91-C95] | 1.8 (1.60-2.01) | 1.43 (1.21-1.69) | 39.1 |

*Calculated as $[\ln(HR_{\text{minimally-adjusted}}) - \ln(HR_{\text{fully-adjusted}})]/\ln(HR_{\text{minimally-adjusted}}) \times 100]$.

Negative values indicate strengthening of an inverse association upon adjustment, i.e., movement away from the null. Large percentage changes (e.g. lung cancer and melanoma) reflect small values of ln(HR) in minimally adjusted models and were interpreted as indicative of substantial attenuation in fully adjusted models rather than absolute measures.

*NA:* For lung adenocarcinoma, percent attenuation was not calculated as the minimally adjusted estimate was close to null; instead, the direction and magnitude were interpreted qualitatively. The attenuation was linked to adjustment for smoking. For oesophageal SCC, percentage attenuation was not calculated as both minimally and fully adjusted estimates were not statistically different from null.

OAC, oesophageal adenocarcinoma; LHTC, lymphatic and haematopoietic cancers; SCC, squamous cell carcinoma

## Appendix Table 4G HRs and 95% CIs comparing risk of cancer in men to women (Table 4.1 with unformatted p-values)

| | Women (%) | Men (%) | Total count | Minimally-adjusted model[a] | | Multivariable-adjusted model[b] | | Multivariable-adjusted model with cancer specific covariates[c] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | p-value | HR (95% CI) | p-value | Additional covariates | HR (95% CI) | p-value |
| **Oral cavity [C00-C14]** | 324 (33.8) | 634 (66.2) | 958 | 2.34 (2.04-2.67) | 2.15e-35 | 2.06 (1.69-2.51) | 4.99e-13 | | 2.06 (1.69-2.51) | 4.98e-13 |
| **Oesophagus [C15]** | 249 (26.0) | 707 (74.0) | 956 | 3.3 (2.86-3.82) | 5.84e-59 | 2.89 (2.36-3.55) | 3.03e-24 | | 2.89 (2.36-3.55) | 3.02e-24 |
| Adenocarcinoma | 110 (16.2) | 571 (83.8) | 681 | 6.01 (4.9-7.37) | 2.46e-66 | 5.47 (4.19-7.14) | 6.66e-36 | Gastro-oesophageal reflux disease (GORD) | 5.45 (4.18-7.12) | 9.02e-36 |
| Squamous cell carcinoma | 128 (52.5) | 116 (47.5) | 244 | 1.07 (0.83-1.37) | 0.622 | 0.81 (0.55-1.2) | 0.297 | | 0.81 (0.55-1.2) | 0.297 |
| **Stomach [C16]** | 203 (31.3) | 446 (68.7) | 649 | 2.55 (2.16-3.01) | 2.44e-28 | 2.34 (1.83-2.97) | 5.75e-12 | | 2.34 (1.83-2.97) | 5.76e-12 |
| Gastric cardia [C16.0] | 62 (21.8) | 223 (78.2) | 285 | 4.17 (3.15-5.53) | 3.03e-23 | 3.7 (2.51-5.44) | 3.19e-11 | Diabetes | 3.65 (2.48-5.38) | 3.20e-11 |

## Appendix Table 4G HRs and 95% CIs comparing risk of cancer in men to women (Table 4.1 with unformatted p-values)

| | Women (%) | Men (%) | Total count | Minimally-adjusted model[a] | | Multivariable-adjusted model[b] | | Multivariable-adjusted model with cancer specific covariates[c] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | p-value | HR (95% CI) | p-value | Additional covariates | HR (95% CI) | p-value |
| Gastric non-cardia [C16.1-C16.6] | 77 (39.9) | 116 (60.1) | 193 | 1.74 (1.31-2.33) | 1.60e-04 | 1.69 (1.09-2.6) | 0.018 | | 1.69 (1.09-2.6) | 0.018 |
| **Colorectum [C18-C20]** | 2316 (43.0) | 3072 (57.0) | 5388 | 1.55 (1.47-1.64) | 2.00e-57 | 1.25 (1.15-1.36) | 1.34e-07 | Diabetes, inflammatory bowel disease, processed meat | 1.22 (1.12-1.33) | 2.55e-06 |
| Colon [C18] | 1667 (46.6) | 1909 (53.4) | 3576 | 1.34 (1.25-1.43) | 3.03e-18 | 1.06 (0.96-1.17) | 0.254 | Diabetes, inflammatory bowel disease, processed meat | 1.04 (0.94-1.15) | 0.456 |
| Rectum [C19-C20] | 663 (36) | 1178 (64) | 1841 | 2.08 (1.89-2.29) | 1.60e-51 | 1.74 (1.51-2.01) | 1.97e-14 | Diabetes, inflammatory bowel disease, processed meat | 1.7 (1.47-1.96) | 5.80e-13 |
| **Anus [C21]** | 127 (65.8) | 66 (34.2) | 193 | 0.62 (0.46-0.83) | 0.001 | 0.45 (0.29-0.71) | 5.24e-04 | Human immunodeficiency virus (HIV), no. of sexual partners | 0.41 (0.26-0.64) | 9.04e-05 |

## Appendix Table 4G HRs and 95% CIs comparing risk of cancer in men to women (Table 4.1 with unformatted p-values)

| | Women (%) | Men (%) | Total count | Minimally-adjusted model[a] | | Multivariable-adjusted model[b] | | Multivariable-adjusted model with cancer specific covariates[c] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | p-value | HR (95% CI) | p-value | Additional covariates | HR (95% CI) | p-value |
| **Liver [C22]** | 194 (33.4) | 387 (66.6) | 581 | 2.34 (1.97-2.79) | 4.09e-22 | 2 (1.55-2.58) | 1.12e-07 | Alcoholic liver disease, non-alcoholic fatty liver disease, gallbladder disease | 1.91 (1.48-2.47) | 7.06e-07 |
| **Gallbladder [C23]** | 74 (63.8) | 42 (36.2) | 116 | 0.64 (0.44-0.94) | 0.023 | 0.56 (0.31-0.99) | 0.046 | Gallbladder disease | 0.55 (0.31-0.97) | 0.0036 |
| **Pancreas [C25]** | 531 (45.9) | 627 (54.1) | 1158 | 1.38 (1.23-1.55) | 3.85e-08 | 1.04 (0.87-1.25) | 0.642 | | 1.04 (0.87-1.25) | 0.642 |
| **Lung [C34]** | 1779 (48.6) | 1885 (51.4) | 3664 | 1.23 (1.16-1.32) | 2.53e-10 | 0.93 (0.84-1.03) | 0.144 | | 0.93 (0.84-1.03) | 0.144 |
| Adenocarcinoma | 825 (54.2) | 697 (45.8) | 1522 | 0.98 (0.89-1.09) | 0.759 | 0.72 (0.62-0.84) | 2.67e-05 | | 0.72 (0.62-0.84) | 2.67e-05 |
| Squamous cell carcinoma | 293 (37.4) | 490 (62.6) | 783 | 1.94 (1.67-2.24) | 4.37e-19 | 1.41 (1.14-1.75) | 0.002 | | 1.41 (1.14-1.75) | 0.002 |
| Small cell carcinoma | 156 (45) | 191 (55) | 347 | 1.42 (1.15-1.75) | 0.001 | 1.19 (0.86-1.64) | 0.289 | | 1.19 (0.86-1.64) | 0.289 |

## Appendix Table 4G HRs and 95% CIs comparing risk of cancer in men to women (Table 4.1 with unformatted p-values)

| | Women (%) | Men (%) | Total count | Minimally-adjusted model[a] | | Multivariable-adjusted model[b] | | Multivariable-adjusted model with cancer specific covariates[c] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | p-value | HR (95% CI) | p-value | Additional covariates | HR (95% CI) | p-value |
| **Melanoma [C43]** | 1335 (49.3) | 1371 (50.7) | 2706 | 1.22 (1.13-1.31) | 3.74e-07 | 0.87 (0.77-0.98) | 0.019 | Use of UV protection, time spent outdoors in the summer, ease of skin tanning | 0.97 (0.86-1.09) | 0.591 |
| **Breast [C50]** | 8737 (99.4) | 55 (0.6) | 8792 | 0.007 (0.006-0.009) | *p<1e-310** | 0.005 (0.004-0.007) | *p<1e-310** | | 0.005 (0.004-0.007) | *p<1e-310** |
| **Kidney [C64-C65]** | 492 (35.7) | 885 (64.3) | 1377 | 2.11 (1.89-2.36) | 2.80e-40 | 1.84 (1.56-2.16) | 3.83e-13 | Hypertension | 1.77 (1.51-2.09) | 8.32e-12 |
| **Bladder [C67]** | 237 (22.2) | 830 (77.8) | 1067 | 4.07 (3.53-4.71) | 6.58e-81 | 3.47 (2.85-4.24) | 1.19e-34 | | 3.47 (2.85-4.24) | 1.18e-34 |
| **Thyroid [C73]** | 292 (73.7) | 104 (26.3) | 396 | 0.43 (0.34-0.53) | 7.83e-14 | 0.35 (0.25-0.48) | 2.11e-10 | Goitre | 0.36 (0.26-0.49) | 3.76e-10 |
| **Lymphatic and haematopoietic cancers [C81-C96]** | 1886 (43.7) | 2431 (56.3) | 4317 | 1.51 (1.42-1.60) | 6.55e-41 | 1.22 (1.11-1.33) | 2.57e-05 | | 1.22 (1.11-1.33) | 2.56e-05 |
| Hodgkin lymphoma [C81] | 63 (47.7) | 69 (52.3) | 132 | 1.31 (0.93-1.84) | 0.122 | 1.29 (0.77-2.17) | 0.326 | | 1.29 (0.77-2.17) | 0.326 |

## Appendix Table 4G HRs and 95% CIs comparing risk of cancer in men to women (Table 4.1 with unformatted p-values)

| | Women (%) | Men (%) | Total count | Minimally-adjusted model[a] | | Multivariable-adjusted model[b] | | Multivariable-adjusted model with cancer specific covariates[c] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | HR (95% CI) | p-value | HR (95% CI) | p-value | Additional covariates | HR (95% CI) | p-value |
| Non-Hodgkin lymphoma [C82-C85] | 931 (46.4) | 1074 (53.6) | 2005 | 1.35 (1.24-1.47) | 2.37e-11 | 1.04 (0.91-1.19) | 0.585 | | 1.04 (0.91-1.19) | 0.585 |
| Multiple myeloma [C90 & C88] | 383 (43.1) | 505 (56.9) | 888 | 1.54 (1.35-1.76) | 2.16e-10 | 1.35 (1.1-1.65) | 0.004 | | 1.35 (1.10-1.65) | 0.004 |
| Leukaemia [C91-C95] | 500 (39.5) | 766 (60.5) | 1266 | 1.8 (1.60-2.01) | 2.78e-24 | 1.43 (1.21-1.69) | 3.86e-05 | | 1.43 (1.21-1.69) | 3.85e-05 |

[a]Minimally-adjusted hazard ratios (95% CIs) stratifying for age, fifths of Townsend deprivation index and region.

[b]Multivariable-adjusted hazard ratios (95% CIs), additionally adjusted for ethnicity, qualifications, height, BMI, smoking (status and intensity), and alcohol.

[c]Multivariable-adjusted hazard ratios (95% CIs) with additional adjustment for cancer specific covariates. Covariate selection has been described in the text.

*p-value was not retrievable further.

**Appendix Table 5A Hazard ratios (HRs) and 95% confidence intervals (CIs) comparing rectum cancer risk [C19-C20] risk by adiposity measures, using data from 249,602 women and 213,163 men in the UK Biobank.**

| Adiposity Trait | Categories | Women<br>HR (95% CI)<br>*653 cases* | Men<br>HR (95% CI)<br>*1,153 cases* |
|---|---|---|---|
| **BMI**<br>(WHO; kg/m$^2$) | 25-<30 kg/m$^2$ | 1.01 (0.85–1.21) | 1.05 (0.91–1.23) |
| | 30-<35 kg/m$^2$ | 0.93 (0.73–1.17) | 1.20 (1.00–1.43) |
| | >35 kg/m$^2$ | 0.72 (0.50–1.02) | 1.07 (0.82–1.41) |
| | *p-trend* | 0.05 | 0.43 |
| | *p-heterogeneity* | 0.26 | 0.21 |
| **BMI**<br>(quartiles; kg/m$^2$) | Q2 | 0.96 (0.77–1.19) | 0.95 (0.80–1.13) |
| | Q3 | 1.04 (0.84–1.29) | 1.10 (0.93–1.30) |
| | Q4 | 0.88 (0.70–1.11) | 1.16 (0.98–1.38) |
| | *p-trend* | 0.44 | 0.03 |
| | *p-heterogeneity* | 0.53 | 0.08 |
| **Body fat percentage**<br>(BFP) | Q2 | 1.09 (0.87–1.36) | 0.94 (0.79–1.13) |
| | Q3 | 1.10 (0.88–1.37) | 1.05 (0.89–1.25) |
| | Q4 | 0.90 (0.71–1.14) | 1.08 (0.91–1.29) |
| | *p-trend* | 0.41 | 0.21 |
| | *p-heterogeneity* | 0.26 | 0.40 |
| **Hip circumference**<br>(HC; cm) | Q2 | 0.79 (0.63–0.97) | 1.00 (0.85–1.18) |
| | Q3 | 0.96 (0.78–1.20) | 1.15 (0.97–1.36) |
| | Q4 | 0.75 (0.60–0.94) | 1.08 (0.91–1.28) |
| | *p-trend* | 0.08 | 0.19 |
| | *p-heterogeneity* | 0.02 | 0.31 |
| **Waist circumference**<br>(WC; cm) | Q2 | 0.95 (0.76–1.18) | 1.19 (1.00–1.43) |
| | Q3 | 0.96 (0.77–1.20) | 1.32 (1.11–1.58) |
| | Q4 | 0.98 (0.78–1.23) | 1.31 (1.10–1.57) |
| | *p-trend* | 0.89 | 0.002 |
| | *p-heterogeneity* | 0.97 | 0.009 |
| **Waist to hip ratio**<br>(WHR) | Q2 | 1.12 (0.89–1.41) | 1.21 (1.01–1.45) |
| | Q3 | 1.18 (0.94–1.49) | 1.28 (1.07–1.53) |
| | Q4 | 1.13 (0.90–1.43) | 1.32 (1.10–1.58) |
| | *p-trend* | 0.26 | 0.003 |
| | *p-heterogeneity* | 0.53 | 0.02 |
| **Waist to height ratio**<br>(WHtR) | Q2 | 1.08 (0.86–1.35) | 1.10 (0.92–1.31) |
| | Q3 | 0.99 (0.79–1.25) | 1.20 (1.00–1.43) |

| | | | |
|---|---|---|---|
| | Q4 | 1.08 (0.85–1.36) | 1.23 (1.03–1.47) |
| | *p-trend* | 0.72 | 0.02 |
| | *p-heterogeneity* | 0.81 | 0.11 |
| **Trunk fat percentage** (TFP) | Q2 | 0.91 (0.73–1.13) | 1.00 (0.84–1.20) |
| | Q3 | 0.96 (0.77–1.20) | 1.09 (0.92–1.29) |
| | Q4 | 0.84 (0.67–1.06) | 1.13 (0.95–1.34) |
| | *p-trend* | 0.22 | 0.11 |
| | *p-heterogeneity* | 0.48 | 0.39 |

All Cox models were adjusted run with one adiposity trait at a time, additionally stratifying the models by age at recruitment, deprivation index, region, and adjusting for qualifications, ethnicity, height, smoking status, alcohol, and prevalent diabetes, irritable bowel disease, and red meat consumption.

*The BMI (WHO) reference category was <25 kg/m$^2$. For all other adiposity traits, the first quartile (Q1) was the reference category.

**Appendix Table 5B Hazard ratios (95% confidence intervals) comparing men with women for risk of oesophageal adenocarcinoma (ICD-10 C15), adjusted for one adiposity trait at a time.**

| Adiposity Trait | Continuous adjustment | | | | Categorical adjustment | | | |
|---|---|---|---|---|---|---|---|---|
| | HR (95% CI) | p-value | Z-score | %Δ lnHR | HR (95% CI) | p-value | Z-score | %Δ lnHR |
| *No adiposity trait adjustment* | *5.55 (4.24-7.25)* | *<0.001* | *12.56* | *0.0* | *5.55 (4.24-7.25)* | *<0.001* | *12.56* | *0.0* |
| BMI* (kg/m$^2$) | 5.26 (4.02–6.89) | <0.001 | 12.10 | 3.1 | 5.12 (3.91–6.71) | <0.001 | 11.87 | 4.7 |
| Body fat percentage (BFP) | 10.03 (7.36–13.66) | <0.001 | 14.61 | -34.5 | 8.35 (6.08–11.47) | <0.001 | 13.11 | -23.8 |
| Hip circumference (HC; cm) | 6.34 (4.83–8.33) | <0.001 | 13.29 | -7.8 | 6.11 (4.65–8.02) | <0.001 | 13.04 | -5.6 |
| Waist circumference (WC; cm) | 4.48 (3.41–5.89) | <0.001 | 10.79 | 12.5 | 4.40 (3.33–5.81) | <0.001 | 10.44 | 13.5 |
| Waist-to-hip ratio (WHR) | 3.33 (2.48–4.48) | <0.001 | 7.97 | 29.8 | 3.54 (2.59–4.83) | <0.001 | 7.94 | 26.2 |
| Waist-to-height ratio (WHtR) | 4.47 (3.41–5.88) | <0.001 | 10.77 | 12.6 | 4.48 (3.40–5.89) | <0.001 | 10.70 | 12.5 |
| Trunk fat percentage (TFP) | 7.48 (5.66–9.88) | <0.001 | 14.18 | -17.4 | 6.65 (5.04–8.78) | <0.001 | 13.38 | -10.6 |

Models assessed sex differences in OAC risk by adiposity trait. 656 cases of OAC were identified (15.9% in women). All Cox models were run with one adiposity trait at a time and additionally stratified by age at recruitment, deprivation index, region, and adjusted for qualifications, ethnicity, height, smoking status, alcohol, and prevalent gastro-oesophageal reflux disease.
*BMI adjustment results are shown for completeness; cancer-specific multivariable BMI results were presented in Chapter 4. Here, they were recalculated owing to a slight reduction in total cancer cases with additional cohort exclusions. The Chapter 4 HR from a minimally adjusted Cox model (stratified by age at recruitment, Townsend deprivation index, and region, and no further adjustments) was 6.01 (95% CI 4.90–7.37; p<0.001).
%Δ ln(HR) calculated relative to the no–adiposity trait adjustment model, as $[\ln(HR_{\text{unadjusted}}) - \ln(HR_{\text{adjusted}})]/\ln(HR_{\text{unadjusted}}) \times 100$. Positive values indicate attenuation toward the null.

**Appendix Table 5C Hazard ratios (95% confidence intervals) comparing men with women for risk of kidney cancer [ICD-10 C64-C65], adjusted for one adiposity trait at a time.**

| Adiposity Trait | Continuous adjustment | | | | Categorical adjustment | | | |
|---|---|---|---|---|---|---|---|---|
| | HR (95% CI) | p-value | Z-score | %Δln(HR) | HR (95% CI) | p-value | Z-score | %Δln(HR) |
| *No adiposity trait adjustment* | *1.82 (1.54-2.15)* | *<0.001* | *7.17* | *0.0* | *1.82 (1.54-2.15)* | *<0.001* | *7.17* | *0.0* |
| BMI* (kg/m$^2$) | 1.75 (1.49–2.07) | <0.001 | 6.65 | 6.4 | 1.74 (1.47–2.05) | <0.001 | 6.51 | 7.4 |
| Body fat percentage (BFP) | 2.49 (2.04–3.03) | <0.001 | 9.03 | -52.5 | 2.43 (1.99–2.97) | <0.001 | 8.71 | -48.5 |
| Hip circumference (HC; cm) | 1.97 (1.67–2.33) | <0.001 | 7.96 | -13.4 | 1.95 (1.65–2.31) | <0.001 | 7.84 | -11.7 |
| Waist circumference (WC; cm) | 1.57 (1.33–1.86) | <0.001 | 5.27 | 24.6 | 1.58 (1.33–1.88) | <0.001 | 5.27 | 23.4 |
| Waist-to-hip ratio (WHR) | 1.26 (1.05–1.52) | 0.015 | 2.42 | 61.4 | 1.36 (1.13–1.65) | 0.001 | 3.17 | 48.5 |
| Waist-to-height ratio (WHtR) | 1.57 (1.33–1.86) | <0.001 | 5.25 | 24.6 | 1.59 (1.34–1.88) | <0.001 | 5.33 | 22.4 |
| Trunk fat percentage (TFP) | 2.12 (1.78–2.53) | <0.001 | 8.42 | -25.8 | 2.11 (1.77–2.52) | <0.001 | 8.38 | -24.9 |

Models assessed sex differences in kidney cancer risk by adiposity trait. 1349 cases of kidney cancer were identified (35.7% in women). All Cox models were run with one adiposity trait at a time, and additionally stratified by age at recruitment, Townsend deprivation index, and region and adjusted for education (qualifications), ethnicity, height, smoking status, alcohol, and prevalent hypertension.
*BMI adjustment results are shown for completeness; cancer-specific multivariable BMI results were presented in Chapter 4. Here, they were recalculated owing to a slight reduction in total cancer cases with additional cohort exclusions. The Chapter 4 HR from a minimally adjusted Cox model (stratified by age at recruitment, Townsend deprivation index, and region, and no further adjustments) was 2.11 (95% CI 1.89-2.36; p<0.001).
%Δln(HR) calculated relative to the no–adiposity trait adjustment model, as $[\ln(HR_{\text{unadjusted}}) - \ln(HR_{\text{adjusted}})]/\ln(HR_{\text{unadjusted}}) \times 100$. Positive values indicate attenuation toward the null.

**Appendix Table 7A: Summary of hazard ratios (HR, 95% CI) comparing men to women for all[a] cancer sites with higher risk in men.**

| | Main cohort analyses | | | Sensitivity analyses (fully adjusted) | | | Adiposity substitution[d] (%Δln(HR)) | Interpretation and research implications |
|---|---|---|---|---|---|---|---|---|
| | Minimally adjusted[b] | Fully adjusted[b] | %Δ lnHR[c] | Never smokers | Alcohol <20g/day | BMI <25 kg/m$^2$ | | |
| Oesophageal adenocarcinoma [C15] | 6.0 (4.9-7.4) | 5.5 (4.2-7.1) | 5.5% | 5.1 (3.3-8.0) | 5.0 (3.6-7.1) | 4.7 (2.6-8.7) | WC: 4.4 (3.3-5.8); 13.5% WHR: 3.5 (2.6-4.8); 26.2% | Central adiposity traits partially contribute to excess risk in men, but there remains a large male predominance. Further work on mechanistic insights and molecular (hormonal and metabolic) risk factors is needed. |
| Gastric cardia [C16.0] | 4.2 (3.2-5.5) | 3.7 (2.5-5.4) | 9.3% | – | 4.0 (2.4-6.6) | – | | The role of smoking and adiposity could not be further evaluated due to insufficient cases in restricted analyses. |
| Bladder [C67] | 4.1 (3.5-4.7) | 3.5 (2.9-4.2) | 11.4% | 3.2 (2.3-4.4) | 3.6 (2.8-4.6) | 4.4 (3.0-6.4) | | Sex differences remain upon evaluation from risk factors. Further work on occupational exposures and genetic risk factors (Y chromosome) is needed. |
| Oral cavity [C00-C14] | 2.3 (2.0-2.7) | 2.1 (1.7-2.5) | 15.0% | 1.8 (1.3-2.5) | 1.5 (1.2-2.0) | 2.4 (1.7-3.3) | | Alcohol consumption may partially contribute to the observed sex difference. As analyses were restricted to light drinkers, there may be residual confounding from alcohol. |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Liver [C22] | 2.3 (2.0-2.8) | 1.9 (1.5-2.5) | 23.9% | 2.4 (1.6-3.6) | 1.8 (1.3-2.6) | 1.7 (0.7-2.0) | | Excess adiposity contributes substantially for the observed sex difference after accounting for major liver-related conditions. |
| Kidney [C64-C65] | 2.1 (1.9-2.4) | 1.8 (1.5-2.1) | 23.5% | 2.1 (1.6-2.7) | 1.7 (1.4-2.0) | 2.0 (1.4-2.8) | WC: 1.6 (1.3-1.9); 23.4% WHR: 1.4 (1.1-1.7); 48.5% | Central adiposity traits strongly contribute to the excess risk in men. Further work on molecular risk factors underpinning the waist-based measures is needed. |
| Rectum [C19-C20] | 2.1 (1.9-2.2) | 1.7 (1.5-2.0) | 27.5% | 1.9 (1.5-2.4) | 1.7 (1.4-2.0) | 1.5 (1.2-2.0) | Not associated with adiposity traits. | Sex differences persist across restrictions. Further work on risk factors specific to rectal cancer is needed. |
| Leukaemia [C91-C95] | 1.8 (1.6-2.0) | 1.4 (1.2-1.7) | 39.1% | 1.4 (1.1-1.9) | 1.4 (1.1-1.7) | 1.5 (1.1-2.1) | | Sex differences were modest and attenuated, with estimates close to the null in restricted cohort analyses. |

[a]Among the 15 cancers sites and 13 subtypes evaluated, these cancer sites had higher incidence in men compared to women in the cancer-site specific multivariable adjusted models (Chapter 4).
[b]Minimally adjusted Cox PH models included stratifying for age at recruitment, deprivation, and region, and cancer- site specific multivariable adjusted models included additionally adjusting for ethnicity, educational qualifications, height, BMI, smoking, alcohol, and cancer site-specific risk factors.
[c]%Δln(HR) from minimally to fully adjusted models. Calculated as $[\ln(HR_{\text{unadjusted}}) - \ln(HR_{\text{adjusted}})]/\ln(HR_{\text{unadjusted}}) \times 100]$. Positive values indicate attenuation toward the null.
[d]Analyses assessed sex differences in OAC and Kidney cancer risk by adiposity trait. Percent change in ln(HR) calculated relative to the no-adiposity trait adjustment model.
Bladder cancer was not followed up for adiposity analyses.
(–) denotes sites <100 cancer cases, for which HRs were not calculated.

**Appendix Table 7B: Summary of hazard ratios (HR, 95% CI) comparing men to women, for all cancer sites with higher risk in women.**

| | Main cohort analyses | | | Sensitivity analyses (fully adjusted) | | | Interpretation and research implications |
|---|---|---|---|---|---|---|---|
| | Minimally adjusted[b] | Fully adjusted[b] | %Δ lnHR[c] | Never smokers | Alcohol <20g/day | BMI <25 kg/m$^2$ | |
| *Breast [C50]* | *0.007 (0.006-0.009)* | *0.005 (0.006-0.009)* | *-6.8%* | *0.005 (0.003-0.007)* | *0.006 (0.004-0.008)* | *0.005 (0.003-0.010)* | *Breast cancer was included in the analyses as a positive control, demonstrating expected and stable sex differences across models.* |
| Thyroid | 0.4 (0.3-0.5) | 0.4 (0.3-0.5) | −20.7% | 0.4 (0.2-0.54) | 0.4 (0.2-0.5) | 0.4 (0.2-0.6) | Sex differences were consistent across restrictions. The observed differences are widely attributable to diagnostic differences. Women are observed to experience earlier and more frequent diagnoses due to higher healthcare seeking behaviours and referrals for gynaecological concerns. |
| Anus | 0.6 (0.5-0.8) | 0.4 (0.3-0.6) | −86.6% | – | 0.3 (0.2-0.5) | – | Sex differences strengthened after adjustment; however, the contribution of smoking and adiposity could not be evaluated due to limited case numbers. Further research with greater statistical power is needed. |
| Lung adenocarcinoma | 0.98 (0.9-1.1) | 0.72 (0.6-0.8) | *NA* | 0.6 (0.4-0.9) | 0.7 (0.6-0.8) | 0.9 (0.9-1.1) | Lung adenocarcinoma in never smokers was followed up separately and assessed for several risk factors additionally, however the observed sex differences remained. Further research on sex differences with greater statistical power and a focus on non-smoking-related risk factors is needed. |

[a]Among the 15 cancers sites and 13 subtypes evaluated, these cancer sites had higher incidence in women compared to men, summarised from Chapter 4 and 6.

[b]Minimally adjusted Cox PH models included stratifying for age at recruitment, deprivation, and region, and fully adjusted models included additionally adjusting for ethnicity, educational qualifications, height, BMI, smoking, alcohol, and cancer site-specific risk factors.

[c]%Δln(HR) from from minimal to fully adjusted models. Calculated as $[\ln(HR_{\text{unadjusted}}) - \ln(HR_{\text{adjusted}})]/\ln(HR_{\text{unadjusted}}) \times 100]$. Negative values indicate strengthening of an inverse association upon adjustment, i.e., movement away from the null.

(*NA*) For lung adenocarcinoma, percent attenuation was not calculated as the minimally adjusted estimate was close to null; instead, the direction and magnitude were interpreted qualitatively. The attenuation was linked to smoking, as adjustment for smoking revealed the higher risk for women.
(–) denotes sites <100 cancer cases, for which HRs were not calculated.

# References


ABNET, C. C., NYREN, O. & ADAMI, H. O. 2018. Esophageal Cancer. *In:* ADAMI, H.-O., HUNTER, D. J., LAGIOU, P. & MUCCI, L. (eds.) *Textbook of Cancer Epidemiology.* 3 ed. New York: Oxford University Press.

ADAMI, H.-O., HUNTER, D. J., LAGIOU, P. & MUCCI, L. 2018. *Textbook of Cancer Epidemiology,* New York, Oxford University Press.

ANDERSON, W. F., JATOI, I., TSE, J. & ROSENBERG, P. S. 2010. Male breast cancer: a population-based comparison with female breast cancer. *J Clin Oncol,* 28**,** 232–9. https://doi.org/10.1200/JCO.2009.23.8162

ARAGHI, M., SOERJOMATARAM, I., BARDOT, A., FERLAY, J., CABASAG, J. C., MORRISON, S. D., et al.ARNOLD, M. 2019. Changes in colorectal cancer incidence in seven high-income countries: a population-based study. *The Lancet Gastroenterology & Hepatology,* 4**,** 511–518. https://doi.org/10.1016/S2468-1253(19)30147-5

ARMITAGE, O. J., GASCOYNE, D. R., LUNNING, A. M. & CAVALLI, F. 2017. Non-Hodgkin lymphoma. *The Lancet,* 390**,** 298–310. https://doi.org/10.1016/s0140-6736(16)32407-2

ARNOLD, M., FERLAY, J., HENEGOUWEN, B. V. I. M. & SOERJOMATARAM, I. 2020. Global burden of oesophageal and gastric cancer by histology and subsite in 2018. *Gut,* 69**,** 1564–1571. https://doi.org/10.1136/gutjnl-2020-321600

ARNOLD, M., LAVERSANNE, M., BROWN, M. L., DEVESA, S. S. & BRAY, F. 2017. Predicting the Future Burden of Esophageal Cancer by Histological Subtype: International Trends in Incidence up to 2030. *American Journal of Gastroenterology,* 112**,** 1247–1255. https://doi.org/10.1038/ajg.2017.155

ARNOLD, M., SOERJOMATARAM, I., FERLAY, J. & FORMAN, D. 2015. Global incidence of oesophageal cancer by histological subtype in 2012. *Gut,* 64**,** 381–387. https://doi.org/10.1136/gutjnl-2014-308124

ARNOLD, P. A. 2017. A general theory of sexual differentiation. *Journal of Neuroscience Research,* 95**,** 291–300. https://doi.org/10.1002/jnr.23884

AVEYARD, P., LEWIS, A., TEARNE, S., HOOD, K., CHRISTIAN-BROWN, A., ADAB, P., et al.JEBB, S. A. 2016. Screening and brief intervention for obesity in primary care: a parallel, two-arm, randomised trial. *The Lancet,* 388**,** 2492–2500. https://doi.org/10.1016/S0140-6736(16)31893-1

BAGNARDI, V., ROTA, M., BOTTERI, E., SCOTTI, L., JENAB, M., BELLOCCO, R., et al.BOFFETTA, P. 2011. Alcohol consumption and lung cancer risk in never smokers: a

meta-analysis. *Annals of Oncology,* 22**,** 2631–2639. https://doi.org/10.1093/annonc/mdr027

BAGNARDI, V., ROTA, M., BOTTERI, E., TRAMACERE, I., ISLAMI, F., FEDIRKO, V., et al.VECCHIA, L. C. 2015. Alcohol consumption and site-specific cancer risk: a comprehensive dose–response meta-analysis. *British Journal of Cancer,* 112**,** 580–593. https://doi.org/10.1038/bjc.2014.579

BAIN, C., FESKANICH, D., SPEIZER, E. F., THUN, M., HERTZMARK, E., ROSNER, A. B. & COLDITZ, A. G. 2004. Lung Cancer Rates in Men and Women With Comparable Histories of Smoking. *JNCI Journal of the National Cancer Institute,* 96**,** 826–834. https://doi.org/10.1093/jnci/djh143

BAMIA, C., STUVER, S. & MUCCI, L. A. 2018. Cancer of the Liver and Biliary Tract. *In:* ADAMI, H.-O., HUNTER, D. J., LAGIOU, P. & MUCCI, L. (eds.) *Textbook of Cancer Epidemiology.* 3 ed. New York: Oxford University Press.

BAUER, G. R., CHURCHILL, S. M., MAHENDRAN, M., WALWYN, C., LIZOTTE, D. & VILLA-RUEDA, A. A. 2021. Intersectionality in quantitative research: A systematic review of its emergence and applications of theory and methods. *SSM - Population Health,* 14**,** 100798. https://doi.org/10.1016/j.ssmph.2021.100798

BEELEN, R., RAASCHOU-NIELSEN, O., STAFOGGIA, M., ANDERSEN, J. Z., WEINMAYR, G., HOFFMANN, B., et al.HOEK, G. 2014. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. *The Lancet,* 383**,** 785–795. https://doi.org/10.1016/s0140-6736(13)62158-3

BELLOTT, D. W., HUGHES, J. F., SKALETSKY, H., BROWN, L. G., PYNTIKOVA, T., CHO, T.-J., et al.PAGE, D. C. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature,* 508**,** 494–499. 10.1038/nature13206

BENIGNI, R. 2007. Social sexual inequality and sex difference in cancer incidence. *Environmental Research,* 104**,** 128–134. https://doi.org/10.1016/j.envres.2006.09.007

BERNARDEZ, B., HIGUERA, O., MARTINEZ-CALLEJO, V., CARDEñA-GUTIéRREZ, A., RODRíGUEZ, M. A. J., BERTRáN, S. A., et al.MORENO-MARTíNEZ, M.-E. 2025. Sex and gender differences in cancer pathogenesis and pharmacology. *Clinical and Translational Oncology,* 27**,** 3837–3848. https://doi.org/10.1007/s12094-025-03894-1

BMJ RESEARCH NEWS 2013. "Women who smoke like men, die like men". *BMJ,* 346**,** f543–f543. https://doi.org/10.1136/bmj.f543

BOFFETTA, P. 2011. I. Epidemiology of adult non-Hodgkin lymphoma. *Annals of Oncology,* 22**,** iv27–iv31. https://doi.org/10.1093/annonc/mdr167

BOURS, J. L. M. 2021. Tutorial: A nontechnical explanation of the counterfactual definition of effect modification and interaction. *Journal of Clinical Epidemiology,* 134**,** 113–124. https://doi.org/10.1016/j.jclinepi.2021.01.022

BOURS, J. L. M. 2023. Using mediators to understand effect modification and interaction. *Journal of Clinical Epidemiology,* 163**,** 117–121. https://doi.org/10.1016/j.jclinepi.2023.09.005

BRANNON, R. A., HAAKE, M. S., HACKER, E. K., PRUTHI, S. R., WALLEN, M. E., NIELSEN, E. M. & RATHMELL, K. W. 2012. Meta-analysis of Clear Cell Renal Cell Carcinoma Gene Expression Defines a Variant Subgroup and Identifies Gender Influences on Tumor Biology. *European Urology,* 61**,** 258–268. https://doi.org/10.1016/j.eururo.2011.10.007

BREDELLA, A. M. 2017. Sex Differences in Body Composition. *Advances in Experimental Medicine and Biology*.

BRENNER, N., MENTZER, J. A., HILL, M., ALMOND, R., ALLEN, N., PAWLITA, M. & WATERBOER, T. 2020. Characterization of human papillomavirus (HPV) 16 E6 seropositive individuals without HPV-associated malignancies after 10 years of follow-up in the UK Biobank. *eBioMedicine,* 62**,** 103123. https://doi.org/10.1016/j.ebiom.2020.103123

BRINTON, A. L., KEY, J. T., KOLONEL, N. L., MICHELS, B. K., SESSO, D. H., URSIN, G., et al.COOK, B. M. 2015. Prediagnostic Sex Steroid Hormones in Relation to Male Breast Cancer Risk. *Journal of Clinical Oncology,* 33**,** 2041–2050. https://doi.org/10.1200/JCO.2014.59.1602

BRINTON, A. L., RICHESSON, A. D., GIERACH, L. G., LACEY, V. J., PARK, Y., HOLLENBECK, R. A. & SCHATZKIN, A. 2008. Prospective Evaluation of Risk Factors for Male Breast Cancer. *JNCI Journal of the National Cancer Institute,* 100**,** 1477–1481. https://doi.org/10.1093/jnci/djn329

BRITTON, A., BEN-SHLOMO, Y., BENZEVAL, M., KUH, D. & BELL, S. 2015. Life course trajectories of alcohol consumption in the United Kingdom using longitudinal data from nine cohort studies. *BMC Medicine,* 13**,** 47. https://doi.org/10.1186/s12916-015-0273-z

BROWN, F. K., RUMGAY, H., DUNLOP, C., RYAN, M., QUARTLY, F., COX, A., et al.PARKIN, M. D. 2018. The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. *British Journal of Cancer,* 118**,** 1130–1141. https://doi.org/10.1038/s41416-018-0029-6

BROWN, J., PFEIFFER, R. M., SHREWSBURY, D., OCALLAGHAN, S., BERNER, A. M., GADALLA, S. M., et al.JACKSON, S. S. 2023. Prevalence of cancer risk factors among transgender and gender diverse individuals: a cross-sectional analysis using UK primary care data. *British Journal of General Practice,* 73**,** e486–e492. https://doi.org/10.3399/BJGP.2023.0023

BRUCE, N., CHOUDHURY, I., DHERANI, M., ADAIR-ROHANI, H. & POPE, D. 2013. CHAPTER 9. HOUSEHOLD AIR POLLUTION. *In:* STRAIF, K., COHEN, A. & SAMET, J. (eds.) *International Agency for Research on Cancer (IARC) Scientific Publication No. 161: Air Pollution and Cancer.* Lyon, France: World Health Organization.

BRUHN-OLSZEWSKA, B., MARKLJUNG, E., RYCHLICKA-BUNIOWSKA, E., SARKISYAN, D., FILIPOWICZ, N. & DUMANSKI, P. J. 2025. The effects of loss of Y chromosome on

male health. *Nature Reviews Genetics,* 26**,** 320–335. https://doi.org/10.1038/s41576-024-00805-y

BURGER, M., CATTO, W. F. J., DALBAGNI, G., GROSSMAN, B. H., HERR, H., KARAKIEWICZ, P., et al.LOTAN, Y. 2013. Epidemiology and Risk Factors of Urothelial Bladder Cancer. *European Urology,* 63**,** 234–241. http://dx.doi.org/10.1016/j.eururo.2012.07.033

BURTON, A., TATARU, D., DRIVER, J. R., BIRD, G. T., HUWS, D., WALLACE, D., et al.AITHAL, G. 2021. Primary liver cancer in the UK: Incidence, incidence-based mortality, and survival by subtype, sex, and nation. *JHEP Reports,* 3**,** 100232. https://doi.org/10.1016/j.jhepr.2021.100232

CANCER RESEARCH UK. 2023. *Cancer risks in the workplace* [Online]. Available: https://www.cancerresearchuk.org/about-cancer/causes-of-cancer/cancer-risks-in-the-workplace.

CANCER RESEARCH UK. 2025. *Cancer Research UK: Cancer Statistics Data Hub* [Online]. London: Cancer Research UK. Available: https://crukcancerintelligence.shinyapps.io/CancerStatsDataHub/.

CANCER RESEARCH UK. 2026. *Worldwide cancer statistics* [Online]. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics/worldwide-cancer.

CENTRE FOR DISEASE CONTROL AND PREVENTION. 2024. *Lung Cancer Among People Who Never Smoked* [Online]. Centre for Disease Control and Prevention. Available: https://www.cdc.gov/lung-cancer/nonsmokers/index.html.

CHADEAU-HYAM, M., BODINIER, B., VERMEULEN, R., KARIMI, M., ZUBER, V., CASTAGNé, R., et al.DELPIERRE, C. 2020. Education, biological ageing, all-cause and cause-specific mortality and morbidity: UK biobank cohort study. *EClinicalMedicine,* 29-30**,** 100658. https://doi.org/10.1016/j.eclinm.2020.100658

CHANDANOS, E. & LAGERGREN, J. 2009. The mystery of male dominance in oesophageal cancer and the potential protective role of oestrogen. *Eur J Cancer,* 45**,** 3149–55. https://doi.org/10.1016/j.ejca.2009.09.001

CHATURVEDI, K. A., FREEDMAN, D. N. & ABNET, C. C. 2022. The Evolving Epidemiology of Oral Cavity and Oropharyngeal Cancers. *Cancer Research,* 82**,** 2821–2823. https://doi.org/10.1158/0008-5472.CAN-22-2124

CHAUDHARY, P., SINGHA, B., ABDEL-HAFIZ, A. H., VELEGRAKI, M., SUNDI, D., SATTURWAR, S., et al.LI, X. 2025. Sex differences in bladder cancer: understanding biological and clinical implications. *Biology of Sex Differences,* 16. https://doi.org/10.1186/s13293-025-00715-6

CHEN, T.-H. D., HIRST, J., COUPLAND, A. C. C., LIAO, W., BALDWIN, R. D. & HIPPISLEY-COX, J. 2025. Ethnic disparities in lung cancer incidence and differences in diagnostic characteristics: a population-based cohort study in England. *The Lancet Regional Health - Europe,* 48**,** 101124. https://doi.org/10.1016/j.lanepe.2024.101124

CHEN, Z., LEE, L., CHEN, J., COLLINS, R., WU, F., GUO, Y., et al.PETO, R. 2005. Cohort Profile: The Kadoorie Study of Chronic Disease in China (KSCDC). *International Journal of Epidemiology,* 34**,** 1243–1249. https://doi.org/10.1093/ije/dyi174

CLAYTON, A. J. & TANNENBAUM, C. 2016. Reporting Sex, Gender, or Both in Clinical Research? *JAMA,* 316**,** 1863. https://doi.org/10.1001/jama.2016.16405

CLOCCHIATTI, A., CORA, E., ZHANG, Y. & DOTTO, P. G. 2016. Sexual dimorphism in cancer. *Nature Reviews Cancer,* 16**,** 330–339. https://doi.org/10.1038/nrc.2016.30

COLDITZ, A. G. & HANKINSON, E. S. 2005. The Nurses' Health Study: lifestyle and health among women. *Nature Reviews Cancer,* 5**,** 388–396. https://doi.org/10.1038/nrc1608

COLLINS, R. 2012. What makes UK Biobank special? *The Lancet,* 379**,** 1173–1174. https://doi.org/10.1016/s0140-6736(12)60404-8

COOK, B. M., WOOD, N. S., CASH, D. B., YOUNG, P., ACOSTA, D. R., FALK, T. R., et al.TAYLOR, R. P. 2015. Association Between Circulating Levels of Sex Steroid Hormones and Barrett's Esophagus in Men: A Case–Control Analysis. *Clinical Gastroenterology and Hepatology,* 13**,** 673–682. http://dx.doi.org/10.1016/j.cgh.2014.08.027

COOK, M. B., ADAMS, N., ADJETEY, A., ARATHIMOS, R., BALABANOVIC, M., BLACKWOOD, R., et al.ALI, R. 2025a. Cohort Profile: Our Future Health. *International Journal of Epidemiology,* 54. https://doi.org/10.1093/ije/dyaf171

COOK, M. B., DAWSEY, S. M., FREEDMAN, N. D., INSKIP, P. D., WICHNER, S. M., QURAISHI, S. M., et al.MCGLYNN, K. A. 2009. Sex disparities in cancer incidence by period and age. *Cancer Epidemiol Biomarkers Prev,* 18**,** 1174–82. https://doi.org/10.1158/1055-9965.EPI-08-1118

COOK, M. B., SANDERSON, S. C., DEANFIELD, J. E., REDDINGTON, F., RODDAM, A., HUNTER, D. J. & ALI, R. 2025b. Our Future Health: a unique global resource for discovery and translational research. *Nature Medicine,* 31**,** 728–730. https://doi.org/10.1038/s41591-024-03438-0

COOK, M. B., WILD, C. P. & FORMAN, D. 2005. A Systematic Review and Meta-Analysis of the Sex Ratio for Barrett's Esophagus, Erosive Reflux Disease, and Nonerosive Reflux Disease. *American Journal of Epidemiology,* 162**,** 1050–1061. 10.1093/aje/kwi325

COOPER, C. S., CROFT, S., DAY, R., THOMSON, S. C. & TRUDGILL, J. N. 2009. Patients with prostate cancer are less likely to develop oesophageal adenocarcinoma: could androgens have a role in the aetiology of oesophageal adenocarcinoma? *Cancer Causes & Control,* 20**,** 1363–1368. https://doi.org/10.1007/s10552-009-9359-2

COUPLAND, H. V., ALLUM, W., BLAZEBY, M. J., MENDALL, A. M., HARDWICK, H. R., LINKLATER, M. K., et al.DAVIES, A. E. 2012. Incidence and survival of oesophageal and gastric cancer in England between 1998 and 2007, a population-based study. *BMC Cancer,* 12**,** 11. https://doi.org/10.1186/1471-2407-12-11

CRAIG, L. C., MARSHALL, L. A., SJöSTRöM, M., BAUMAN, E. A., BOOTH, L. M., AINSWORTH, E. B., et al.OJA, P. 2003. International Physical Activity Questionnaire: 12-Country Reliability and Validity. *Medicine & Science in Sports & Exercise,* 35**,** 1381–1395. https://doi.org/10.1249/01.mss.0000078924.61453.fb

CRENSHAW, K. W. 1989. Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicago Legal Forum,* 1989. http://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8

CUMBERBATCH, K. G. M., JUBBER, I., BLACK, C. P., ESPERTO, F., FIGUEROA, D. J., KAMAT, M. A., et al.CATTO, W. F. J. 2018. Epidemiology of Bladder Cancer: A Systematic Review and Contemporary Update of Risk Factors in 2018. *European Urology,* 74**,** 784–795. https://doi.org/10.1016/j.eururo.2018.09.001

DALTVEIT, S. D., MORGAN, E., COLOMBET, M., STELIAROVA-FOUCHER, E., BENDAHHOU, K., MARCOS-GRAGERA, R., et al.SOERJOMATARAM, I. 2025. Global patterns of leukemia by subtype, age, and sex in 185 countries in 2022. *Leukemia,* 39**,** 412–419. https://doi.org/10.1038/s41375-024-02452-y

DARBY, S., HILL, D., AUVINEN, A., BARROS-DIOS, M. J., BAYSSON, H., BOCHICCHIO, F., et al.DOLL, R. 2005. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *BMJ,* 330**,** 223. https://doi:10.1136/bmj.38308.477650.63

DAUGHERTY, E. S., LACEY, V. JAMES, PFEIFFER, M. R., PARK, Y., HOOVER, N. R. & SILVERMAN, T. D. 2013. Reproductive factors and menopausal hormone therapy and bladder cancer risk in the NIH-AARP Diet and Health Study. *International Journal of Cancer,* 133**,** 462–472. https://doi.org/10.1002/ijc.28022

DE BLOK, C. J. M., WIEPJES, C. M., NOTA, N. M., ENGELEN, K. V., ADANK, M. A., DREIJERINK, K. M. A., et al.HEIJER, M. D. 2019. Breast cancer risk in transgender people receiving hormone treatment: nationwide cohort study in the Netherlands. *BMJ***,** l1652. https://doi.org/10.1136/bmj.l1652

DE NIE, I., BLOK, C. J. M. D., SLUIS, T. M. V. D., BARBé, E., PIGOT, G. L. S., WIEPJES, C. M., et al.HEIJER, M. D. 2020. Prostate Cancer Incidence under Androgen Deprivation: Nationwide Cohort Study in Trans Women Receiving Hormone Treatment. *The Journal of Clinical Endocrinology & Metabolism,* 105**,** e3293–e3299. https://doi.org/10.1210/clinem/dgaa412

DEMB, J., EARLES, A., MARTíNEZ, E. M., BUSTAMANTE, R., BRYANT, K. A., MURPHY, D. J., et al.GUPTA, S. 2019. Risk factors for colorectal cancer significantly vary by anatomic site. *BMJ Open Gastroenterology,* 6**,** e000313. https://doi.org/10.1136/bmjgast-2019-000313

DENOS, M., SUN, Y.-Q., BRUMPTON, M. B., LI, Y., ALBANES, D., BURNETT-HARTMAN, A., et al.MAI, X.-M. 2024. Sex hormones and risk of lung and colorectal cancers in women: a Mendelian randomization study. *Scientific Reports,* 14. https://doi.org/10.1038/s41598-024-75305-4

DIRICAN, D. C. & NELSON, S. P. 2024. Y Chromosome Loss and Implications for Oncology. *Molecular Cancer Research,* 22**,** 603–612. https://doi.org/10.1158/1541-7786.MCR-24-0105

DOLL, R. & HILL, B. A. 1954. The Mortality of Doctors in Relation to Their Smoking Habits. *BMJ,* 1**,** 1451–1455. https://doi.org/10.1136/bmj.1.4877.1451

DOLL, R. & PETO, R. 1981. The Causes of Cancer: Quantitative Estimates of Avoidable Risks of Cancer in the United States Today. *JNCI: Journal of the National Cancer Institute,* 66**,** 1192–1308. https://doi.org/10.1093/jnci/66.6.1192

DOLL, R., PETO, R., WHEATLEY, K., GRAY, R. & SUTHERLAND, I. 1994. Mortality in relation to smoking: 40 years' observations on male British doctors. *BMJ,* 309**,** 901–911. https://doi.org/10.1136/bmj.309.6959.901

DONG, M., CIOFFI, G., WANG, J., WAITE, K. A., OSTROM, Q. T., KRUCHKO, C., et al.BARNHOLTZ-SLOAN, J. S. 2020. Sex Differences in Cancer Incidence and Survival: A Pan-Cancer Analysis. *Cancer Epidemiol Biomarkers Prev,* 29**,** 1389–1397. https://doi.org/10.1158/1055-9965.EPI-20-0036

DUNFORD, A., WEINSTOCK, M. D., SAVOVA, V., SCHUMACHER, E. S., CLEARY, P. J., YODA, A., et al.LANE, A. A. 2017. Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. *Nature Genetics,* 49**,** 10–16. https://doi.org/10.1038/ng.3726

EDGREN, G., LIANG, L., ADAMI, H. O. & CHANG, E. T. 2012. Enigmatic sex disparities in cancer incidence. *Eur J Epidemiol,* 27**,** 187–96. https://doi.org/10.1007/s10654-011-9647-5

ELBASHEER, A. M. M., BOHRMANN, B., CHEN, Y., LV, J., SUN, D., WU, X., et al.YANG, L. 2024. Reproductive factors and risk of lung cancer among 300,000 Chinese female never-smokers: evidence from the China Kadoorie Biobank study. *BMC Cancer,* 24. https://doi.org/10.1186/s12885-024-12133-9

ELBASHEER, M. 2024. *Education and cancer incidence and mortality: a prospective study of 0.5 million adults in the China Kadoorie Biobank.* DPhil in Population Health, University of Oxford.

ERVIK, M., LAM, F., LAVERSANNE, M., COLOMBET, M., FERLAY, J., MIRANDA-FILHO, A. & BRAY, F. 2024. Cancer Over Time [Internet database] [Online]. Available: https://gco.iarc.fr/overtime/en/dataviz/trends.

EUROPEAN COMMISSION. 2025. *Gender equality in research and innovation [Internet]* [Online]. Luxembourg: European Commission. Available: https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/democracy-and-rights/gender-equality-research-and-innovation_en.

FENG, Q., BEšEVIć, J., CONROY, M., OMIYALE, W., LACEY, B. & ALLEN, N. 2024. Comparison of body composition measures assessed by bioelectrical impedance analysis versus dual-energy X-ray absorptiometry in the United Kingdom Biobank. *Clinical Nutrition ESPEN,* 63**,** 214–225. https://doi.org/10.1016/j.clnesp.2024.06.040

FERLAY, J., ERVIK, M., LAM, F., LAVERSANNE, M., COLOMBET, M., MERY, L., et al.BRAY, F. 2024. *Global Cancer Observatory: Cancer Today* [Online]. Lyon, France: International Agency for Research on Cancer. Available: https://gco.iarc.who.int/today.

FERNIHOUGH, A. & MCGOVERN, E. M. 2015. Physical stature decline and the health status of the elderly population in England. *Economics & Human Biology,* 16**,** 30–44. https://doi.org/10.1016/j.ehb.2013.12.010

FINK, H., LANGSELIUS, O., VIGNAT, J., RUMGAY, H., REHM, J., MARTINEZ, R. X., et al.SOERJOMATARAM, I. 2026. Global and regional cancer burden attributable to modifiable risk factors to inform prevention. *Nature Medicine,* 32**,** 1306–1315. https://doi.org/10.1038/s41591-026-04219-7

FISHER, J. A., LIAO, L. M., GRAUBARD, B. I., KAUFMAN, J. D., SILVERMAN, D. T. & JONES, R. R. 2025. Ambient air pollution exposure and lung cancer risk in a large prospective U.S. cohort. *Science of The Total Environment,* 1002**,** 180490. 10.1016/j.scitotenv.2025.180490

FORSBERG, A. L., RASI, C., MALMQVIST, N., DAVIES, H., PASUPULATI, S., PAKALAPATI, G., et al.DUMANSKI, P. J. 2014. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nature Genetics,* 46**,** 624–628. https://doi.org/10.1038/ng.2966

FREEDMAN, D. N., LEITZMANN, F. M., HOLLENBECK, R. A., SCHATZKIN, A. & ABNET, C. C. 2008. Cigarette smoking and subsequent risk of lung cancer in men and women: analysis of a prospective cohort study. *The Lancet Oncology,* 9**,** 649–656. https://doi.org/10.1016/S1470-2045(08)70154-2

FRY, A., LITTLEJOHNS, T. J., SUDLOW, C., DOHERTY, N., ADAMSKA, L., SPROSEN, T., et al.ALLEN, N. E. 2017. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol,* 186**,** 1026–1034. https://doi.org/10.1093/aje/kwx246

FU, B. C., SONG, M., LI, X., HAN, J., ADAMI, H. O., GIOVANNUCCI, E. L. & MUCCI, L. A. 2020a. Height as a mediator of sex differences in cancer risk. *Annals of Oncology,* 31**,** 634–640. https://doi.org/10.1016/j.annonc.2020.02.010

FU, B. C., SONG, M., LI, X., HAN, J., ADAMI, H. O., GIOVANNUCCI, E. L. & MUCCI, L. A. 2020b. Height as a mediator of sex differences in cancer risk. *Ann Oncol,* 31**,** 634–640. https://doi.org/10.1016/j.annonc.2020.02.010

GAETA, A., TAGLIABUE, M., D'ECCLESIIS, O., GHIANI, L., MAUGERI, P., BERARDINIS, D. R., et al.CHIOCCA, S. 2023. Are sex and gender considered in head and neck cancer clinical studies? *npj Precision Oncology,* 7. https://doi.org/10.1038/s41698-023-00439-z

GAO, Y., ZHANG, J., ZHAO, H., GUAN, F. & ZENG, P. 2021. Instrumental Heterogeneity in Sex-Specific Two-Sample Mendelian Randomization: Empirical Results From the Relationship Between Anthropometric Traits and Breast/Prostate Cancer. *Frontiers in Genetics,* 12. https://doi.org/10.3389/fgene.2021.651332

GATENBY, C. A. P., HAINSWORTH, A., CAYGILL, C., WATSON, A. & WINSLET, M. 2011. Projections for oesophageal cancer incidence in England to 2033. *European Journal of Cancer Prevention,* 20**,** 283–286. https://doi.org/10.1097/cej.0b013e32834572d2

GILLISON, L. M., CHATURVEDI, K. A., ANDERSON, F. W. & FAKHRY, C. 2015. Epidemiology of Human Papillomavirus–Positive Head and Neck Squamous Cell Carcinoma. *Journal of Clinical Oncology,* 33**,** 3235–3242. https://doi.org/10.1200/jco.2015.61.6995

GIOVANNUCCI, E., HARLAN, M. D., ARCHER, C. M., BERGENSTAL, M. R., GAPSTUR, M. S., HABEL, A. L., et al.YEE, D. 2010. Diabetes and Cancer. *Diabetes Care,* 33**,** 1674–1685. https://doi.org/10.2337/dc10-0666

GOODMAN, S. N. & SAMET, J. M. 2017. Causal Inference in Cancer Epidemiology. *In:* THUN, M., LINET, M. S., CERHAN, J. R., HAIMAN, C. A. & SCHOTTENFELD, D. (eds.) *Cancer Epidemiology and Prevention.* 4 ed. New York: Oxford Scholarship Online.

GREEN, A. C., OLSEN, C. M. & HUNTER, D. J. 2018. Skin Cancer. *In:* ADAMI, H.-O., HUNTER, D. J., LAGIOU, P. & MUCCI, L. (eds.) *Textbook of Cancer Epidemiology.* 3 ed. New York: Oxford University Press.

GREEN, J., CAIRNS, J. B., CASABONNE, D., WRIGHT, L. F., REEVES, G. & BERAL, V. 2011. Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *The Lancet Oncology,* 12**,** 785–794. https://doi.org/10.1016/S1470-2045(11)70154-1

GREEN, J., RODDAM, A., PIRIE, K., KIRICHEK, O., REEVES, G. & BERAL, V. 2012. Reproductive factors and risk of oesophageal and gastric cancer in the Million Women Study cohort. *British Journal of Cancer,* 106**,** 210–216. https://doi.org/10.1038/bjc.2011.525

GRIFFIN, A. B., ANDERSON, L. G., SHIH, A. R. & WHITSEL, A. E. 2012. Use of alternative time scales in Cox proportional hazard models: implications for time-varying environmental exposures. *Statistics in Medicine,* 31**,** 3320–3327. https://doi.org/10.1002/sim.5347

GRULICH, A. E., JIN, F. & POYNTEN, M. 2017. Anal Cancer. *In:* THUN, M., LINET, M. S., CERHAN, J. R., HAIMAN, C. A. & SCHOTTENFELD, D. (eds.) *Cancer Epidemiology and Prevention.* 4 ed.: Oxford Academic.

GUNNELL, D., OKASHA, M., SMITH, D. G., OLIVER, E. S., SANDHU, J. & HOLLY, P. M. J. 2001. Height, Leg Length, and Cancer Risk: A Systematic Review. *Epidemiologic Reviews,* 23**,** 313–342. https://doi.org/10.1093/oxfordjournals.epirev.a000809

HADKHALE, K., MARTINSEN, I. J., WEIDERPASS, E., KJæRHEIM, K., SPARéN, P., TRYGGVADóTTIR, L., et al.PUKKALA, E. 2019. Occupational variation in bladder cancer in Nordic males adjusted with approximated smoking prevalence. *Acta Oncologica,* 58**,** 29–37. https://doi.org/10.1080/0284186x.2018.1518591

HAMRA, G. B., GUHA, N., COHEN, A., LADEN, F., RAASCHOU-NIELSEN, O., SAMET, J. M., et al.LOOMIS, D. 2014. Outdoor Particulate Matter Exposure and Lung Cancer: A Systematic Review and Meta-Analysis. *Environmental Health Perspectives,* 122**,** 906–911. 10.1289/ehp/1408092

HAUPT, S., CARAMIA, F., KLEIN, S. L., RUBIN, J. B. & HAUPT, Y. 2021. Sex disparities matter in cancer development and therapy. *Nat Rev Cancer,* 21**,** 393–407. https://doi.org/10.1038/s41568-021-00348-y

HEALTH AND SOCIAL CARE INFORMATION CENTRE 2013. Statistics on Smoking: England, 2013. Leeds.Available: https://extranet.who.int/fctcapps/sites/default/files/2023-04/uk_2016_annex2_smoke_2013.pdf.

HEIDARI, S., BABOR, F. T., CASTRO, D. P., TORT, S. & CURNO, M. 2016. Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use. *Research Integrity and Peer Review,* 1. https://doi.org/10.1186/s41073-016-0007-6

HENNEKENS, H. C. & EBERLEIN, K. 1985. A randomized trial of aspirin and β-carotene among U.S. physicians. *Preventive Medicine,* 14**,** 165–168. https://doi.org/10.1016/0091-7435(85)90031-3

HERNAN, M. A. & ROBINS, J. M. 2025. *Causal Inference: What If?,* Boca Raton, Chapman & Hall/CRC.

HIATT, A. R., XUN, W., SANTIAGO-RODRíGUEZ, J. E., PIKHARTOVA, J. & SHELTON, N. 2025. Socio-economic position, area-level deprivation and gradients in cancer incidence: England and Wales, 1971–2016. *BMC Public Health,* 25. https://doi.org/10.1186/s12889-025-21875-5

HILL, B. A. 1965. The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine,* 58**,** 295–300. https://doi.org/10.1177/003591576505800503

HILL, W., LIM, E. L., WEEDEN, C. E., LEE, C., AUGUSTINE, M., CHEN, K., et al.SWANTON, C. 2023. Lung adenocarcinoma promotion by air pollutants. *Nature,* 616**,** 159–167. 10.1038/s41586-023-05874-3

HJALGRIM, H. & JARRETT, F. R. 2020. Epidemiology of Hodgkin Lymphoma. *In:* ENGERT, A. & YOUNES, A. (eds.) *Hodgkin Lymphoma: A Comprehensive Overview.* Cham, Switzerland: Springer Nature Switzerland AG.

HODGE, M. J., PATEL, V. A., ISLAMI, F., JEMAL, A. & HIATT, A. R. 2023. Educational Attainment and Cancer Incidence in a Large Nationwide Prospective Cohort. *Cancer Epidemiology, Biomarkers & Prevention,* 32**,** 1747–1755. https://doi.org/10.1158/1055-9965.EPI-23-0290

HOSGOOD, D. H., BOFFETTA, P., GREENLAND, S., LEE, A. Y.-C., MCLAUGHLIN, J., SEOW, A., et al.LAN, Q. 2010. In-Home Coal and Wood Use and Lung Cancer Risk: A Pooled Analysis of the International Lung Cancer Consortium. *Environmental Health Perspectives,* 118**,** 1743–1747. https://doi.org/10.1289/ehp.1002217

HOSGOOD, D. H., SONG, M., HSIUNG, A. C., YIN, Z., SHU, X.-O., WANG, Z., et al.LAN, Q. 2015. Interactions between household air pollution and GWAS-identified lung cancer susceptibility markers in the Female Lung Cancer Consortium in Asia (FLCCA). *Human Genetics,* 134**,** 333–341. 10.1007/s00439-014-1528-z

HOUSE OF COMMONS LIBRARY 2025. Obesity statistics: Research Briefing Number 03336 [Internet]. London: PARLIAMENT, U.Available: https://researchbriefings.files.parliament.uk/documents/SN03336/SN03336.pdf.

HSIEH, F. Y. & LAVORI, P. W. 2000. Sample-Size Calculations for the Cox Proportional Hazards Regression Model with Nonbinary Covariates. *Controlled Clinical Trials,* 21**,** 552–560. https://doi.org/10.1016/S0197-2456(00)00104-5

HUANG, J., CHAN, C. S., LOK, V., ZHANG, L., LUCERO-PRISNO, E. D., XU, W., et al.WONG, S. C. M. 2022a. The epidemiological landscape of multiple myeloma: a global cancer registry estimate of disease burden, risk factors, and temporal trends. *The Lancet Haematology,* 9**,** e670–e677. https://doi.org/10.1016/S2352-3026(22)00165-X

HUANG, J., DENG, Y., TIN, S. M., LOK, V., NGAI, H. C., ZHANG, L., et al.WONG, C. S. M. 2022b. Distribution, Risk Factors, and Temporal Trends for Lung Cancer Incidence and Mortality. *Chest,* 161**,** 1101–1111. https://doi.org/10.1016/j.chest.2021.12.655

HUANG, Y., ZHU, M., JI, M., FAN, J., XIE, J., WEI, X., et al.SHEN, H. 2021. Air Pollution, Genetic Factors, and the Risk of Lung Cancer: A Prospective Study in the UK Biobank. *American Journal of Respiratory and Critical Care Medicine,* 204**,** 817–825. https://doi.org/10.1164/rccm.202011-4063oc

HUXLEY, R. R., ANSARY-MOGHADDAM, A., CLIFTON, P., CZERNICHOW, S., PARR, L. C. & WOODWARD, M. 2009. The impact of dietary and lifestyle risk factors on risk of colorectal cancer: A quantitative overview of the epidemiological evidence. *International Journal of Cancer,* 125**,** 171–180. https://doi.org/10.1002/ijc.24343

HVIDTFELDT, A. U., SEVERI, G., ANDERSEN, J. Z., ATKINSON, R., BAUWELINCK, M., BELLANDER, T., et al.FECHT, D. 2021. Long-term low-level ambient air pollution exposure and risk of lung cancer – A pooled analysis of 7 European cohorts. *Environment International,* 146**,** 106249. https://doi.org/10.1016/j.envint.2020.106249

IARC WORKING GROUP ON THE EVALUATION OF CARCINOGENIC RISKS TO HUMANS 2012. *Personal Habits and Indoor Combustions: Tobacco Smoking,* Lyon, France, International Agency for Research on Cancer.

ILIC, M. & ILIC, I. 2022. Epidemiology of stomach cancer. *World Journal of Gastroenterology,* 28**,** 1187–1203. https://doi.org/10.3748/wjg.v28.i12.1187

INAMURA, K. 2017. Lung Cancer: Understanding Its Molecular Pathology and the 2015 WHO Classification. *Frontiers in Oncology,* 7. https://doi.org/10.3389/fonc.2017.00193

INSTITUTE OF MEDICINE (US) 1994. *Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies: Volume I,* Washington (DC), National Academies Press (US).

INSTITUTE OF MEDICINE (US) 2001. *Exploring the Biological Contributions to Human Health: Does Sex Matter?,* Washington (DC), National Academies Press (US).

INTERNATIONAL AGENCY FOR RESEARCH ON CANCER 1987. *Overall evaluations of carcinogenicity: An updating of IARC Monographs volumes 1–42,* Lyon, France, International Agency for Research on Cancer.

INTERNATIONAL AGENCY FOR RESEARCH ON CANCER 2013. IARC: Outdoor air pollution a leading environmental cause of cancer deaths. 221 ed. Lyon, France: International Agency for Research on Cancer.

INTERNATIONAL AGENCY FOR RESEARCH ON CANCER 2017. New IARC report: comprehensive analysis of factors driving obesity. Lyon, France: CANCER, I. A. F. R. O.Available: https://www.iarc.who.int/wp-content/uploads/2018/07/pr253_E.pdf.

INTERNATIONAL AGENCY FOR RESEARCH ON CANCER 2025a. Global Cancer Observatory [Online]. Available: https://gco.iarc.fr/en.

INTERNATIONAL AGENCY FOR RESEARCH ON CANCER 2025b. List of classifications by cancer sites with *sufficient* or *limited evidence* in humans. Lyon, France.Available: https://monographs.iarc.who.int/wp-content/uploads/2019/07/Classifications_by_cancer_site.pdf.

JACKSON, E. S., BEARD, E., KUJAWSKI, B., SUNYER, E., MICHIE, S., SHAHAB, L., et al.BROWN, J. 2019. Comparison of Trends in Self-reported Cigarette Consumption and Sales in England, 2011 to 2018. *JAMA Network Open,* 2**,** e1910161. https://10.1001/jamanetworkopen.2019.10161

JACKSON, S. S., ADAMI, H.-O., ANDREOTTI, G., BEANE-FREEMAN, E. L., GONZáLEZ, D. B. A., BURING, E. J., et al.KOSHIOL, J. 2020. Associations between reproductive factors and biliary tract cancers in women from the Biliary Tract Cancers Pooling Project. *Journal of Hepatology,* 73**,** 863–872. https://doi.org/10.1016/j.jhep.2020.04.046

JACKSON, S. S., MARKS, M. A., KATKI, H. A., COOK, M. B., HYUN, N., FREEDMAN, N. D., et al.CHATURVEDI, A. K. 2022. Sex disparities in the incidence of 21 cancer types: Quantification of the contribution of risk factors. *Cancer,* 128**,** 3531–3540. https://10.1002/cncr.34390

JEMAL, A., MILLER, D. K., MA, J., SIEGEL, L. R., FEDEWA, A. S., ISLAMI, F., et al.THUN, J. M. 2018. Higher Lung Cancer Incidence in Young Women Than Young Men in the United States. *New England Journal of Medicine,* 378**,** 1999–2009. https://doi.org/10.1056/NEJMoa1715907

JEMAL, A., SCHAFER, J. E., STAR, J., BANDI, P., SUNG, H., ISLAMI, F. & SIEGEL, L. R. 2025. Lung cancer incidence rates in young women and men by state in the United States. *International Journal of Cancer,* 156**,** 499–504. https://doi.org/10.1002/ijc.35188

JEON, J., CAO, P. & MEZA, R. 2025. Lung cancer incidence trends by histology and individual- and county-level sociodemographic characteristics in the United States from 2000 to 2019. *JNCI Monographs,* 2025**,** 211–223. 10.1093/jncimonographs/lgaf003

JHA, P., RAMASUNDARAHETTIGE, C., LANDSMAN, V., ROSTRON, B., THUN, M., ANDERSON, N. R., et al.PETO, R. 2013. 21st-Century Hazards of Smoking and Benefits of Cessation in the United States. *New England Journal of Medicine,* 368**,** 341–350. https://doi.org/10.1056/NEJMsa1211128

JUBBER, I., ONG, S., BUKAVINA, L., BLACK, C. P., COMPéRAT, E., KAMAT, M. A., et al.CUMBERBATCH, G. M. 2023. Epidemiology of Bladder Cancer in 2023: A Systematic Review of Risk Factors. *European Urology,* 84**,** 176–190. https://doi.org/10.1016/j.eururo.2023.03.029

KABAT, C. G., SILVERA, N. A. S., MILLER, B. A. & ROHAN, E. T. 2007. A cohort study of reproductive and hormonal factors and renal cell cancer risk in women. *British Journal of Cancer,* 96**,** 845–849. https://doi.org/10.1038/sj.bjc.6603629

KABAT, G. C., KIM, M. Y., HOLLENBECK, A. R. & ROHAN, T. E. 2014. Attained height, sex, and risk of cancer at different anatomic sites in the NIH-AARP diet and health study. *Cancer Causes Control,* 25**,** 1697–706. https://doi.org/10.1007/s10552-014-0476-1

KAISINGER, R. L., KENTISTOU, A. K., STANKOVIC, S., GARDNER, J. E., DAY, R. F., ZHAO, Y., et al.PERRY, R. B. J. 2023. Large-scale exome sequence analysis identifies sex- and age-specific determinants of obesity. *Cell Genomics,* 3**,** 100362. https://doi.org/10.1016/j.xgen.2023.100362

KAPLAN, J. B. & BENNETT, T. 2003. Use of Race and Ethnicity in Biomedical Publication. *JAMA,* 289**,** 2709. 10.1001/jama.289.20.2709

KENFIELD, A. S., WEI, K. E., STAMPFER, J. M., ROSNER, A. B. & COLDITZ, A. G. 2008. Comparison of aspects of smoking among the four histological types of lung cancer. *Tobacco Control,* 17**,** 198–204. https://doi.org/10.1136/tc.2007.022582

KEUM, N., SONG, M., GIOVANNUCCI, L. E. & ELIASSEN, H. A. 2017. *Obesity and Body Composition*.

KEUM, N., WU, K., GIOVANNUCCI, E. L. & HUNTER, D. J. 2018. Colorectal Cancer. *In:* ADAMI, H.-O., HUNTER, D. J., LAGIOU, P. & MUCCI, L. (eds.) *Textbook of Cancer Epidemiology.* 3 ed. New York: Oxford University Press.

KHAN, M., PAPIER, K., PIRIE, L. K., KEY, J. T., ATKINS, J. & TRAVIS, C. R. 2025. Sex differences in cancer incidence: prospective analyses in the UK Biobank. *British Journal of Cancer,* 133**,** 216–226. https://doi.org/10.1038/s41416-025-03028-y

KIM, W. Y., JOO, D.-H., KIM, Y. S., PARK, S. Y., JANG, S., LEE, H. J., et al.LEE, C.-T. 2025. Gender Disparities and Lung Cancer Screening Outcomes Among Individuals Who Have Never Smoked. *JAMA Network Open,* 8**,** e2454057. https://10.1001/jamanetworkopen.2024.54057

KIRKALI, Z., CHAN, T., MANOHARAN, M., ALGABA, F., BUSCH, C., CHENG, L., et al.WEIDER, J. 2005. Bladder cancer: Epidemiology, staging and grading, and diagnosis. *Urology,* 66**,** 4–34. https://doi.org/10.1016/j.urology.2005.07.062

KJAERHEIM, K., HALDORSEN, T., LYNGE, E., MARTINSEN, J. I., PUKKALA, E., WEIDERPASS, E. & GRIMSRUD, T. K. 2018. Variation in Nordic Work-Related Cancer Risks after Adjustment for Alcohol and Tobacco. *International Journal of Environmental Research and Public Health,* 15**,** 2760. 10.3390/ijerph15122760

KLEIN, L. S. & FLANAGAN, L. K. 2016. Sex differences in immune responses. *Nature Reviews Immunology,* 16**,** 626–638. https://doi.org/10.1038/nri.2016.90

KOEHLER, C., THIELEN, S., GINZKEY, C., HACKENBERG, S., SCHERZED, A., BURGHARTZ, M., et al.KLEINSASSER, N. H. 2013. Nitrogen dioxide is genotoxic in urban concentrations. *Inhalation Toxicology,* 25**,** 341–347. 10.3109/08958378.2013.788104

KOGEVINAS, M., FIGUEROA, D. J., GARCIA-CLOSAS, M. & MUCCI, L. A. 2018. Urinary Bladder Cancer. *In:* ADAMI, H.-O., HUNTER, D. J., LAGIOU, P. & MUCCI, L. (eds.) *Textbook of Cancer Epidemiology.* 3 ed. New York: Oxford University Press.

KOLE, M. T., MUISER, S., KRAFT, M., SIDDIQUI, S., FABBRI, M. L., RABE, F. K., et al.BERGE, D. V. M. 2024. Sex differences in asthma control, lung function and exacerbations: the ATLANTIS study. *BMJ Open Respiratory Research,* 11**,** e002316. https://doi.org/10.1136/bmjresp-2024-002316

KOUTROS, S., KIEMENEY, A. L., CHOUDHURY, P. P., MILNE, L. R., MATURANA, D. L. E., YE, Y., et al.ROTHMAN, N. 2023. Genome-wide Association Study of Bladder Cancer Reveals New Biological and Translational Insights. *European Urology,* 84**,** 127–137. https://doi.org/10.1016/j.eururo.2023.04.020

KRATZER, B. T., BANDI, P., FREEDMAN, D. N., SMITH, A. R., TRAVIS, D. W., JEMAL, A. & SIEGEL, L. R. 2024. Lung cancer statistics, 2023. *Cancer,* 130**,** 1330–1348. https://doi.org/10.1002/cncr.35128

KUKHAREVA, V. P., CAVERLY, J. T., LI, H., KATKI, A. H., CHEUNG, C. L., REESE, J. T., et al.KAWAMOTO, K. 2022. Inaccuracies in electronic health records smoking data and a potential approach to address resulting underestimation in determining lung cancer screening eligibility. *Journal of the American Medical Informatics Association,* 29**,** 779–788. https://doi.org/10.1093/jamia/ocac020

LAGERGREN, J. & NYRéN, O. 1998. Do sex hormones play a role in the etiology of esophageal adenocarcinoma? A new hypothesis tested in a population-based cohort of prostate cancer patients. *Cancer Epidemiology, Biomarkers & Prevention,* 9. https://pubmed.ncbi.nlm.nih.gov/9796637/

LAGIOU, P., TRICHOPOULOS, D. & ADAMI, H. O. 2018. Concepts in Cancer Epidemiology and Etiology. *In:* ADAMI, H.-O., HUNTER, D. J., LAGIOU, P. & MUCCI, L. (eds.) *Textbook of Cancer Epidemiology.* 3 ed. New York: Oxford University Press.

LAKSHMIKANTH, T., CONSIGLIO, C., SARDH, F., FORLIN, R., WANG, J., TAN, Z., et al.BRODIN, P. 2024. Immune system adaptation during gender-affirming testosterone treatment. *Nature,* 633**,** 155–164. https://doi.org/10.1038/s41586-024-07789-z

LAM, J., ALDRIDGE, R., BLACKBURN, R. & HARRON, K. 2023. How is ethnicity reported, described, and analysed in health research in the UK? A bibliographical review and focus group discussions with young refugees. *BMC Public Health,* 23. 10.1186/s12889-023-16947-3

LAN, Q., HSIUNG, A. C., MATSUO, K., HONG, Y.-C., SEOW, A., WANG, Z., et al.ROTHMAN, N. 2012. Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nature Genetics,* 44**,** 1330–1335. https://doi.org/10.1038/ng.2456

LASKAR, S. R., MULLER, C. D., LI, P., MACHIELA, J. M., YE, Y., GABORIEAU, V., et al.SCELO, G. 2019. Sex specific associations in genome wide association analysis of renal cell carcinoma. *European Journal of Human Genetics,* 27**,** 1589–1598. https://doi.org/10.1038/s41431-019-0455-9

LASKAR, S. R., MURPHY, N., FERRARI, P., BRENNAN, P., CROSS, J. A., GUEVARA, M., et al.BORCH, B. K. 2025. A prospective investigation of early-onset colorectal cancer risk factors–pooled analysis of three large-scale European cohorts. *British Journal of Cancer*. https://doi.org/10.1038/s41416-025-03303-y

LAUBY-SECRETAN, B., SCOCCIANTI, C., LOOMIS, D., GROSSE, Y., BIANCHINI, F. & STRAIF, K. 2016. Body Fatness and Cancer — Viewpoint of the IARC Working Group. *New England Journal of Medicine,* 375**,** 794–798. https://10.1056/NEJMsr1606602

LAVORANDO, E., OWENS, M. C. & LIU, K. F. 2024. Comparing the roles of sex chromosome-encoded protein homologs in gene regulation. *Genes & Development*. 10.1101/gad.351890.124

LECLAIR, K., BELL, L. J. K., FURUYA-KANAMORI, L., DOI, A. S., FRANCIS, O. D. & DAVIES, L. 2021. Evaluation of Gender Inequity in Thyroid Cancer Diagnosis. *JAMA Internal Medicine,* 181**,** 1351. https://doi.org/10.1001/jamainternmed.2021.4804

LEEMAN, D., ARNOLD, S., CLARKE, E., COSTELLA, A., IJAZ, S., LYONS, S., et al.SIMMONS, R. 2024. Hepatitis B in England 2024. UK HEALTH SECURITY AGENCY, U. G.Available: https://www.gov.uk/government/publications/hepatitis-b-in-england/hepatitis-b-in-england-2024#foreword.

LEMIEUX, S., PRUD'HOMME, D., BOUCHARD, C., TREMBLAY, A. & DESPRéS, J. 1993. Sex differences in the relation of visceral adipose tissue accumulation to total body fatness. *The American Journal of Clinical Nutrition,* 58**,** 463–467. https://doi.org/10.1093/ajcn/58.4.463

LEPAGE, C., RACHET, B., JOOSTE, V., FAIVRE, J. & COLEMAN, P. M. 2008. Continuing Rapid Increase in Esophageal Adenocarcinoma in England and Wales. *The American Journal of Gastroenterology,* 103**,** 2694–2699. https://doi.org/10.1111/j.1572-0241.2008.02191.x

LER, D. L., GHOSH, S., CHAI, X., THIKE, A. A., HENG, L. H., SIEW, Y. E., et al.TEH, T. B. 2017. Loss of tumor suppressor KDM6A amplifies PRC2-regulated transcriptional repression in bladder cancer and can be targeted through inhibition of EZH2. *Science Translational Medicine,* 9**,** eaai8312. 10.1126/scitranslmed.aai8312

LETT, E., ASABOR, E., BELTRáN, S., CANNON, A. M. & ARAH, O. A. 2022. Conceptualizing, Contextualizing, and Operationalizing Race in Quantitative Health Sciences Research. *The Annals of Family Medicine,* 20**,** 157–163. https://doi.org/10.1370/afm.2792

LI, H. & DURBIN, R. 2024. Genome assembly in the telomere-to-telomere era. *Nature Reviews Genetics,* 25**,** 658–670. https://doi.org/10.1038/s41576-024-00718-w

LI, H. C., PROKOPEC, D. S., SUN, X. R., YOUSIF, F., SCHMITZ, N., AL-SHAHROUR, F., et al.MERING, V. C. 2020. Sex differences in oncogenic mutational processes. *Nature Communications,* 11. https://doi.org/10.1038/s41467-020-17359-2

LI, Y., XU, A., JIA, S. & HUANG, J. 2019. Recent advances in the molecular mechanism of sex disparity in hepatocellular carcinoma (Review). *Oncology Letters*. https://doi.org/10.3892/ol.2019.10127

LIN, H. J. & GIOVANNUCCI, E. 2010. Sex Hormones and Colorectal Cancer: What Have We Learned So Far? *JNCI Journal of the National Cancer Institute,* 102**,** 1746–1747. https://doi.org/10.1093/jnci/djq444

LINDBLAD, M., RODRíGUEZ, G. A. L., CHANDANOS, E. & LAGERGREN, J. 2006. Hormone replacement therapy and risks of oesophageal and gastric adenocarcinomas. *British Journal of Cancer,* 94**,** 136–141. https://doi.org/10.1038/sj.bjc.6602906

LISSOWSKA, J., BARDIN-MIKOLAJCZAK, A., FLETCHER, T., ZARIDZE, D., SZESZENIA-DABROWSKA, N., RUDNAI, P., et al.BOFFETTA, P. 2005. Lung Cancer and Indoor Pollution from Heating and Cooking with Solid Fuels. *American Journal of Epidemiology,* 162**,** 326–333. https://doi.org/10.1093/aje/kwi204

LIU, B., FLOUD, S., YANG, L., YANG, O. T., DU, H., YAO, P., et al.REEVES, K. G. 2025. Prospective associations of diabetes with 15 cancers in 2.2 million UK and Chinese adults. *JNCI: Journal of the National Cancer Institute,* 117**,** 2477–2487. https://doi.org/10.1093/jnci/djaf154

LONARDO, A., NASCIMBENI, F., BALLESTRI, S., FAIRWEATHER, D., WIN, S., THAN, A. T., et al.SUZUKI, A. 2019. Sex Differences in Nonalcoholic Fatty Liver Disease: State of the Art and Identification of Research Gaps. *Hepatology,* 70**,** 1457–1469. https://doi.org/10.1002/hep.30626

LOOMIS, D., HUANG, W. & CHEN, G. 2014. The International Agency for Research on Cancer (IARC) evaluation of the carcinogenicity of outdoor air pollution: focus on China. *Chinese Journal of Cancer,* 33**,** 189–196. https://doi.org/10.5732/cjc.014.10028

LOPES-RAMOS, M. C., QUACKENBUSH, J. & DEMEO, L. D. 2020. Genome-Wide Sex and Gender Differences in Cancer. *Frontiers in Oncology,* 10. https://doi.org/10.3389/fonc.2020.597788

LOPICCOLO, J., GUSEV, A., CHRISTIANI, C. D. & JäNNE, A. P. 2024. Lung cancer in patients who have never smoked — an emerging disease. *Nature Reviews Clinical Oncology,* 21**,** 121–146. https://doi.org/10.1038/s41571-023-00844-0

LOUIS, N. D., PERRY, A., WESSELING, P., BRAT, J. D., CREE, A. I., FIGARELLA-BRANGER, D., et al.ELLISON, W. D. 2021. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro-Oncology,* 23**,** 1231–1251. https://doi.org/10.1093/neuonc/noab106

LUCCA, I., KLATTE, T., FAJKOVIC, H., MARTINO, D. M. & SHARIAT, F. S. 2015. Gender differences in incidence and outcomes of urothelial and kidney cancer. *Nature Reviews Urology,* 12**,** 585–592. https://doi.org/10.1038/nrurol.2015.232

LYRATZOPOULOS, G., ABEL, A. G., MCPHAIL, S., NEAL, D. R. & RUBIN, P. G. 2013. Gender inequalities in the promptness of diagnosis of bladder and renal cancer after symptomatic presentation: evidence from secondary analysis of an English primary care audit survey. *BMJ Open,* 3**,** e002861. https://doi.org/10.1136/bmjopen-2013-002861

MACINNIS, J. R., ENGLISH, R. D., HOPPER, L. J. & GILES, G. G. 2006. Body size and composition and the risk of gastric and oesophageal adenocarcinoma. *International Journal of Cancer,* 118**,** 2628–2631. https://doi.org/10.1002/ijc.21638

MACMAHON, B., PUGH, T. F. & IPSEN, J. 1960. *Epidemiologic Methods,* Boston, Little, Brown and Company.

MALCOMSON, C. F., PARRA-SOTO, S., LU, L., HO, F., CELIS-MORALES, C., SHARP, L. & MATHERS, C. J. 2024. Socio-demographic variation in adherence to the World Cancer Research Fund (WCRF)/American Institute for Cancer Research (AICR) Cancer Prevention Recommendations within the UK Biobank prospective cohort study. *Journal of Public Health,* 46**,** 61–71. https://doi.org/10.1093/pubmed/fdad218

MALHOTRA, J., BOFFETTA, P. & MUCCI, L. A. 2018. Cancer of the Lung, Larynx, and Pleura. *In:* ADAMI, H.-O., HUNTER, D. J., LAGIOU, P. & MUCCI, L. (eds.) *Textbook of Cancer Epidemiology.* 3 ed. New York: Oxford University Press.

MANCINI, F. R., CANO-SANCHO, G., GAMBARETTI, J., MARCHAND, P., BOUTRON-RUAULT, M. C., SEVERI, G., et al.KVASKOFF, M. 2020. Perfluorinated alkylated substances serum concentration and breast cancer risk: Evidence from a nested case-control study in the French E3N cohort. *International Journal of Cancer,* 146**,** 917–928. 10.1002/ijc.32357

MATTEIS, S. D., JARVIS, D., HUTCHINGS, S., DARNTON, A., FISHWICK, D., SADHRA, S., et al.CULLINAN, P. 2016. Occupations associated with COPD risk in the large population-based UK Biobank cohort study. *Occupational and Environmental Medicine,* 73**,** 378–384. 10.1136/oemed-2015-103406

MAUVAIS-JARVIS, F., MERZ, B. N., BARNES, J. P., BRINTON, D. R., CARRERO, J.-J., DEMEO, L. D., et al.SUZUKI, A. 2020. Sex and gender: modifiers of health, disease, and medicine. *The Lancet,* 396**,** 565–582. https://doi.org/10.1016/S0140-6736(20)31561-0

MCGRATH, M., MICHAUD, S. D. & VIVO, D. I. 2006. Hormonal and Reproductive Factors and the Risk of Bladder Cancer in Women. *American Journal of Epidemiology,* 163**,** 236–244. https://doi.org/10.1093/aje/kwj028

MCKEE, S. A., SMITH, P. H., KAUFMAN, M., MAZURE, C. M. & WEINBERGER, A. H. 2016. Sex Differences in Varenicline Efficacy for Smoking Cessation: A Meta-Analysis. *Nicotine & Tobacco Research,* 18**,** 1002–1011. https://doi.org/10.1093/ntr/ntv207

MCMENAMIN, C. Ú., LIU, P., KUNZMANN, T. A., COOK, B. M., COLEMAN, G. H., JOHNSTON, T. B., et al.CARDWELL, R. C. 2021. Circulating Sex Hormones Are Associated With Gastric and Colorectal Cancers but Not Esophageal Adenocarcinoma in the UK Biobank. *American Journal of Gastroenterology,* 116**,** 522–529. https://doi/10.14309/ajg.0000000000001045

MEDEROS, N., FRIEDLAENDER, A., PETERS, S. & ADDEO, A. 2020. Gender-specific aspects of epidemiology, molecular genetics and outcome: lung cancer. *ESMO Open,* 5**,** e000796. https://doi.org/10.1136/esmoopen-2020-000796

MERKATZ, B. R. 1998. Inclusion of Women in Clinical Trials: A Historical Overview of Scientific Ethical and Legal Issues. *Journal of Obstetric, Gynecologic & Neonatal Nursing,* 27**,** 78–84. https://doi.org/10.1111/j.1552-6909.1998.tb02594.x

MIGA, H. K., KOREN, S., RHIE, A., VOLLGER, R. M., GERSHMAN, A., BZIKADZE, A., et al.PHILLIPPY, M. A. 2020. Telomere-to-telomere assembly of a complete human X chromosome. *Nature,* 585**,** 79–84. https://doi.org/10.1038/s41586-020-2547-7

MILLION WOMEN STUDY COLLABORATIVE GROUP 1999. The Million Women Study: design and characteristics of the study population. *Breast Cancer Research,* 1. https://doi.org/10.1186/bcr16

MIRANDA-FILHO, A., PIñEROS, M., FERLAY, J., SOERJOMATARAM, I., MONNEREAU, A. & BRAY, F. 2018. Epidemiological patterns of leukaemia in 184 countries: a population-based study. *The Lancet Haematology,* 5**,** e14–e24. https://doi.org/10.1016/S2352-3026(17)30232-6

MIRANDA-FILHO, A., PIñEROS, M., ZNAOR, A., MARCOS-GRAGERA, R., STELIAROVA-FOUCHER, E. & BRAY, F. 2019. Global patterns and trends in the incidence of non-Hodgkin lymphoma. *Cancer Causes & Control,* 30**,** 489–499. https://doi.org/10.1007/s10552-019-01155-5

MURPHY, C., PANDYA, T., SWANTON, C. & SOLOMON, J. B. 2025. Lung Cancer in Nonsmoking Individuals. *JAMA,* 334**,** 1836. https://doi.org/10.1001/jama.2025.17695

MURPHY, G., DEVESA, S. S., CROSS, A. J., INSKIP, P. D., MCGLYNN, K. A. & COOK, M. B. 2011. Sex disparities in colorectal cancer incidence by anatomic subsite, race and age. *International Journal of Cancer,* 128**,** 1668–1675. 10.1002/ijc.25481

MURPHY, N., JENAB, M. & GUNTER, J. M. 2018. Adiposity and gastrointestinal cancers: epidemiology, mechanisms and future directions. *Nature Reviews Gastroenterology & Hepatology,* 15**,** 659–670. https://doi.org/10.1038/s41575-018-0038-1

MURPHY, N., XU, L., ZERVOUDAKIS, A., XUE, X., KABAT, G., ROHAN, E. T., et al.GUNTER, J. M. 2017. Reproductive and menstrual factors and colorectal cancer incidence in the Women's Health Initiative Observational Study. *British Journal of Cancer,* 116**,** 117–125. https://doi.org/10.1038/bjc.2016.345

NADALETO, B. F., HERBELLA, F. A. & PATTI, M. G. 2016. Gastroesophageal reflux disease in the obese: Pathophysiology and treatment. *Surgery,* 159**,** 475–86. 10.1016/j.surg.2015.04.034

NAM, E. G., CHO, H. K., HAN, K., KIM, M. C., HAN, B., CHO, J. S., et al.PARK, G. Y. 2019. Obesity, abdominal obesity and subsequent risk of kidney cancer: a cohort study of 23.3 million East Asians. *British Journal of Cancer,* 121**,** 271–277. https://doi.org/10.1038/s41416-019-0500-z

NäSLUND, O., STRAND, S. P., SOLHEIM, O., MASRI, A. M., RAPI, O., THURIN, E. & JAKOLA, S. A. 2023. Incidence, management, and outcome of incidental meningioma: what has happened in 10 years? *Journal of Neuro-Oncology,* 165**,** 291–299. https://doi.org/10.1007/s11060-023-04482-5

NATIONAL CANCER INSTITUTE 2025. SEER*Explorer: Cancer statistics interactive application [Online]. Available: https://seer.cancer.gov/statistics-network/explorer.

NATIONAL DISEASE REGISTRATION SERVICE 2021. Children, teenagers and young adults UK cancer statistics report 2021. Leeds, UK.Available: https://digital.nhs.uk/ndrs/data/data-outputs/cancer-publications-and-tools/ctya-uk-cancer-statistics-report-2021

NATIONAL DISEASE REGISTRATION SERVICE 2025. Incidence and mortality dashboard [Online]. Available: https://nhsd-ndrs.shinyapps.io/incidence_and_mortality/.

NATIONAL HEALTH SERVICES. 2024. *Alcohol units* [Online]. National Health Service. Available: https://www.nhs.uk/live-well/alcohol-advice/calculating-alcohol-units/.

NATIONAL HUMAN GENOME RESEARCH INSTITUTE. 2024. *X Chromosome Infographic* [Online]. Available: https://www.genome.gov/about-genomics/fact-sheets/X-Chromosome-facts.

NATIONAL INSTITUTE FOR HEALTH AND CARE RESEARCH. 2020. *Improving inclusion of under-served groups in clinical research: Guidance from the INCLUDE project [Internet]* [Online]. UK: NIHR. Available: https://www.nihr.ac.uk/improving-inclusion-under-served-groups-clinical-research-guidance-include-project.

NATIONAL INSTITUTE FOR HEALTH AND CARE RESEARCH. 2025. *Sex and gender in research policy [Internet]* [Online]. Leeds (UK): NIHR. Available: https://www.nihr.ac.uk/about-us/who-we-are/policies-and-guidelines/sex-and-gender-research.

NELSON, E. D., MOWERY, P., TOMAR, S., MARCUS, S., GIOVINO, G. & ZHAO, L. 2006. Trends in Smokeless Tobacco Use Among Adults and Adolescents in the United States. *American Journal of Public Health,* 96**,** 897–905. https://doi.org/10.2105/AJPH.2004.061580

NHS DIGITAL. 2019. *Statistics on Smoking, England - 2019* [Online]. NHS Digital. Available: https://digital.nhs.uk/data-and-information/publications/statistical/statistics-on-smoking/statistics-on-smoking-england-2019/part-3-smoking-patterns-in-adults-copy.

NHS DIGITAL. 2021. *Cancer Registration Statistics, England 2019* [Online]. Available: https://digital.nhs.uk/data-and-information/publications/statistical/cancer-registration-statistics/england-2019/there-are-higher-rates-of-cancer-in-the-most-deprived-areas.

NHS DIGITAL. 2022a. *Health Survey for England, 2021 part 1: Data tables (HSE 2021 Overweight and obesity tables)* [Online]. Available: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2021/health-survey-for-england-2021-data-tables.

NHS DIGITAL. 2022b. *Health Survey for England, 2021: Part 4 — Trends [Internet]* [Online]. London: NHS Digital. Available: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2021/part-4-trends.

NHS DIGITAL. 2024. *Health Survey for England 2022 Part 2: Adult overweight and obesity* [Online]. London: NHS Digital. Available: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2022-part-2/adult-overweight-and-obesity#:~:text=overweight%20and%20obesity-,Height%20and%20weight,to%2072.8kg%20among%20women.

O'DOHERTY, G. M., FREEDMAN, D. N., HOLLENBECK, R. A., SCHATZKIN, A. & ABNET, C. C. 2012. A prospective cohort study of obesity and risk of oesophageal and gastric adenocarcinoma in the NIH–AARP Diet and Health Study. *Gut,* 61**,** 1261–1268. https://doi.org/10.1136/gutjnl-2011-300551

O'GRADY, J. T., RINALDI, S., MICHELS, A. K., ADAMI, H.-O., BURING, E. J., CHEN, Y., et al.KITAHARA, M. C. 2024. Association of hormonal and reproductive factors with differentiated thyroid cancer risk in women: a pooled prospective cohort analysis. *International Journal of Epidemiology,* 53. https://doi.org/10.1093/ije/dyad172

OFFICE FOR HEALTH IMPROVEMENT & DISPARITIES. 2024. *Obesity profile: short statistical commentary, May 2024 [Internet]* [Online]. London: UK Government. Available: https://www.gov.uk/government/statistics/update-to-the-obesity-profile-on-fingertips/obesity-profile-short-statistical-commentary-may-2024.

OFFICE FOR NATIONAL STATISTICS. 2019. *What is the difference between sex and gender?* [Online]. Newport, UK: Office for National Statistics. Available: https://backup.ons.gov.uk/wp-content/uploads/sites/3/2019/02/What-is-the-difference-between-sex-and-gender_-1.pdf.

OFFICE OF NATIONAL STATISTICS. 2024. *Adult smoking habits in the UK: 2023* [Online]. Office of National Statistics. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/adultsmokinghabitsingreatbritain/2023.

OFFICE OF RESEARCH ON WOMEN'S HEALTH. 2024. *History of women's participation in clinical research [Internet]* [Online]. Bethesda (MD): National Institutes of Health. Available: https://orwh.od.nih.gov/toolkit/recruitment/history.

OFFICE OF RESEARCH ON WOMEN'S HEALTH. 2025. *Sex as a Biological Variable [Internet]* [Online]. Bethesda (MD): National Institutes of Health. Available: https://orwh.od.nih.gov/sex-as-a-biological-variable.

OLSEN, C. M. 2020. Global trends in melanoma mortality differ by sex and age. *British Journal of Dermatology,* 183**,** 985–986. https://doi.org/10.1111/bjd.19166

ORTEGA, S. L., BRADBURY, E. K., CROSS, J. A., MORRIS, S. J., GUNTER, J. M. & MURPHY, N. 2017. A Prospective Investigation of Body Size, Body Fat Composition and Colorectal Cancer Risk in the UK Biobank. *Scientific Reports,* 7. https://doi.org/10.1038/s41598-017-17997-5

OWENS, C. M., YANAS, A. & LIU, F. K. 2024. Sex chromosome-encoded protein homologs: current progress and open questions. *Nature Structural & Molecular Biology,* 31**,** 1156–1166. https://doi.org/10.1038/s41594-024-01362-y

PAPIER, K., FENSOM, K. G., KNUPPEL, A., APPLEBY, N. P., TONG, N. Y. T., SCHMIDT, A. J., et al.PEREZ-CORNAGO, A. 2021. Meat consumption and risk of 25 common conditions: outcome-wide analyses in 475,000 men and women in the UK Biobank study. *BMC Medicine,* 19. https://doi.org/10.1186/s12916-021-01922-9

PARKIN, M. D. 2011. Cancers attributable to infection in the UK in 2010. *British Journal of Cancer,* 105**,** S49–S56. https://doi.org/10.1038/bjc.2011.484

PARRA-SOTO, S., COWLEY, S. E., REZENDE, M. F. L., FERRECCIO, C., MATHERS, C. J., PELL, P. J., et al.CELIS-MORALES, C. 2021. Associations of six adiposity-related markers with incidence and mortality from 24 cancers—findings from the UK Biobank prospective cohort study. *BMC Medicine,* 19. https://doi.org/10.1186/s12916-020-01848-8

PDQ ADULT TREATMENT EDITORIAL BOARD. 2025. *PDQ Non-Small Cell Lung Cancer Treatment* [Online]. Bethesda, MD: National Cancer Institute. Available: https://www.cancer.gov/types/lung/hp/non-small-cell-lung-treatment-pdq.

PEIRED, J. A., CAMPI, R., ANGELOTTI, L. M., ANTONELLI, G., CONTE, C., LAZZERI, E., et al.ROMAGNANI, P. 2021. Sex and Gender Differences in Kidney Cancer: Clinical and Experimental Evidence. *Cancers,* 13**,** 4588. https://doi.org/10.3390/cancers13184588

PEREZ-CORNAGO, A., KEY, J. T., ALLEN, E. N., FENSOM, K. G., BRADBURY, E. K., MARTIN, M. R. & TRAVIS, C. R. 2017. Prospective investigation of risk factors for prostate cancer in the UK Biobank cohort study. *British Journal of Cancer,* 117**,** 1562–1571. https://doi.org/10.1038/bjc.2017.312

PETERS, E. A. S., HUXLEY, R. R. & WOODWARD, M. 2014. Do smoking habits differ between women and men in contemporary Western populations? Evidence from half a million people in the UK Biobank study. *BMJ Open,* 4**,** e005663. http://dx.doi.org/10.1136/bmjopen-2014-005663

PETERSON, E., DE, P. & NUTTALL, R. 2012. BMI, Diet and Female Reproductive Factors as Risks for Thyroid Cancer: A Systematic Review. *PLOS ONE,* 7**,** e29177. https://doi.org/10.1371/journal.pone.0029177

PETO, R. 2023. The need for Ignorance in cancer research. *European Journal of Epidemiology,* 38**,** 1239–1243. https://doi.org/10.1007/s10654-023-01085-2

PETO, R., DARBY, S., DEO, H., SILCOCKS, P., WHITLEY, E. & DOLL, R. 2000. Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. *BMJ,* 321**,** 323–329. https://doi.org/10.1136/bmj.321.7257.323

PETO, R., PIKE, C. M., ARMITAGE, P., BRESLOW, E. N., COX, R. D., HOWARD, V. S., et al.SMITH, G. P. 1976. Design and analysis of randomized clinical trials requiring

prolonged observation of each patient. I. Introduction and design. *British Journal of Cancer,* 34**,** 585–612. https://doi.org/10.1038/bjc.1976.220

PETRICK, L. J. & COOK, B. M. 2020. Do Sex Hormones Underlie Sex Differences in Cancer Incidence? Testing the Intuitive in Esophageal Adenocarcinoma. *American Journal of Gastroenterology,* 115**,** 211–213. https://doi.org/10.14309/ajg.0000000000000511

PETRICK, L. J., HYLAND, L. P., CARON, P., FALK, T. R., PFEIFFER, M. R., DAWSEY, M. S., et al.COOK, B. M. 2019. Associations Between Prediagnostic Concentrations of Circulating Sex Steroid Hormones and Esophageal/Gastric Cardia Adenocarcinoma Among Men. *JNCI: Journal of the National Cancer Institute,* 111**,** 34–41. https://doi.org/10.1093/jnci/djy082

PHYSICIANS' HEALTH STUDY 1989. Final Report on the Aspirin Component of the Ongoing Physicians' Health Study. *New England Journal of Medicine,* 321**,** 129–135. https://doi.org/10.1056/NEJM198907203210301

PINSKY, P. F., CHURCH, T. R., IZMIRLIAN, G. & KRAMER, B. S. 2013. The National Lung Screening Trial: Results stratified by demographics, smoking history, and lung cancer histology. *Cancer,* 119**,** 3976–3983. 10.1002/cncr.28326

PIRIE, K., PETO, R., GREEN, J., REEVES, K. G. & BERAL, V. 2016. Lung cancer in never smokers in the UK Million Women Study. *International Journal of Cancer,* 139**,** 347–354. https://doi.org/10.1002/ijc.30084

PISCHON, T., LAHMANN, H. P., BOEING, H., TJøNNELAND, A., HALKJæR, J., OVERVAD, K., et al.RIBOLI, E. 2006. Body size and risk of renal cell carcinoma in the European Prospective Investigation into Cancer and Nutrition (EPIC). *International Journal of Cancer,* 118**,** 728–738. https://doi.org/10.1002/ijc.21398

POSIT TEAM 2025. RStudio: Integrated Development Environment for R, Posit Software. Boston, MA: PBC.

POWER, L. M. & SCHULKIN, J. 2008. Sex differences in fat storage, fat metabolism, and the health risks from obesity: possible evolutionary origins. *British Journal of Nutrition,* 99**,** 931–940. https://doi.org/10.1017/S0007114507853347

PUETT, R. C., HART, J. E., YANOSKY, J. D., SPIEGELMAN, D., WANG, M., FISHER, J. A., et al.LADEN, F. 2014. Particulate Matter Air Pollution Exposure, Distance to Road, and Incident Lung Cancer in the Nurses' Health Study Cohort. *Environmental Health Perspectives,* 122**,** 926–932. 10.1289/ehp.1307490

PUKKALA, E., MARTINSEN, J. I., LYNGE, E., GUNNARSDOTTIR, H. K., SPARéN, P., TRYGGVADOTTIR, L., et al.KJAERHEIM, K. 2009. Occupation and cancer – follow-up of 15 million people in five Nordic countries. *Acta Oncologica,* 48**,** 646–790. 10.1080/02841860902913546

PULIT, L. S., STONEMAN, C., MORRIS, P. A., WOOD, R. A., GLASTONBURY, A. C., TYRRELL, J., et al.LINDGREN, M. C. 2019. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Human Molecular Genetics,* 28**,** 166–174. https://doi.org/10.1093/hmg/ddy327

PURDUE, P. M., DUTTA, D., MACHIELA, J. M., GORMAN, R. B., WINTER, T., OKUHARA, D., et al.CHANOCK, J. S. 2024. Multi-ancestry genome-wide association study of kidney cancer identifies 63 susceptibility regions. *Nature Genetics,* 56**,** 809–818. https://doi.org/10.1038/s41588-024-01725-7

RADKIEWICZ, C., JOHANSSON, A. L. V., DICKMAN, P. W., LAMBE, M. & EDGREN, G. 2017. Sex differences in cancer risk and survival: A Swedish cohort study. *Eur J Cancer,* 84**,** 130–140. https://doi.org/10.1016/j.ejca.2017.07.013

RAGAVAN, M. & PATEL, I. M. 2022. The evolving landscape of sex-based differences in lung cancer: a distinct disease in women. *European Respiratory Review,* 31**,** 210100. https://doi.org/10.1183/16000617.0100-2021

RASK-ANDERSEN, M., IVANSSON, E., HOGLUND, J., EK, W. E., KARLSSON, T. & JOHANSSON, A. 2023. Adiposity and sex-specific cancer risk. *Cancer Cell,* 41**,** 1186–1197 e4. https://doi.org/10.1016/j.ccell.2023.05.010

RASK-ANDERSEN, M., KARLSSON, T., EK, E. W. & JOHANSSON, Å. 2019. Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nature Communications,* 10. https://doi.org/10.1038/s41467-018-08000-4

RAWLA, P., SUNKARA, T. & GADUPUTI, V. 2019. Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. *World Journal of Oncology,* 10**,** 10–27. https://doi.org/10.14740/wjon1166

RENEHAN, G. A. & SOERJOMATARAM, I. 2016. Obesity as an Avoidable Cause of Cancer (Attributable Risks). *Obesity and Cancer.*

RHIE, A., NURK, S., CECHOVA, M., HOYT, J. S., TAYLOR, J. D., ALTEMOSE, N., et al.PHILLIPPY, M. A. 2023. The complete sequence of a human Y chromosome. *Nature,* 621**,** 344–354. https://doi.org/10.1038/s41586-023-06457-y

RIBOLI, E. 1997. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. *International Journal of Epidemiology,* 26**,** 6S–14. https://doi.org/10.1093/ije/26.suppl_1.s6

ROA, C. J., GARCíA, P., KAPOOR, K. V., MAITHEL, K. S., JAVLE, M. & KOSHIOL, J. 2022. Gallbladder cancer. *Nature Reviews Disease Primers,* 8. https://doi.org/10.1038/s41572-022-00398-y

ROJAS, G. M., POZZI, E. & RAMASAMY, R. 2024. Unlocking the mystery of the human Y chromosome. *Nature Reviews Urology,* 21**,** 65–66. https://doi.org/10.1038/s41585-023-00826-y

ROMAN, E., SMITH, A. R. & MUCCI, L. A. 2018. Leukaemia. *In:* ADAMI, H.-O., HUNTER, D. J., LAGIOU, P. & MUCCI, L. (eds.) *Textbook of Cancer Epidemiology.* 3 ed. New York: Oxford University Press.

ROMERO-CORRAL, A., SOMERS, K. V., SIERRA-JOHNSON, J., THOMAS, J. R., COLLAZO-CLAVELL, L. M., KORINEK, J., et al.LOPEZ-JIMENEZ, F. 2008. Accuracy of body mass

index in diagnosing obesity in the adult general population. *International Journal of Obesity,* 32**,** 959–966. https://doi.org/10.1038/ijo.2008.11

ROY, S., BANERJEE, D., BANERJEE, A. & CHOUDHURY, L. 2025. Sex hormones and gender differences in immune responses and anticancer immunity: a comprehensive review. *Biologia Futura*. https://doi.org/10.1007/s42977-025-00291-1

RUBIN, B. J. 2022. The spectrum of sex differences in cancer. *Trends in Cancer,* 8**,** 303–315. https://doi.org/10.1016/j.trecan.2022.01.013

RUHL, J., CALLAGHAN, C. & SCHUSSLER, N. 2022. *Summary Stage 2018: Codes and Coding Instructions* [Online]. Bethesda, MD. Available: https://seer.cancer.gov/tools/ssm/.

RUSHTON, L., HUTCHINGS, J. S., FORTUNATO, L., YOUNG, C., EVANS, S. G., BROWN, T., et al.TONGEREN, V. M. 2012. Occupational cancer burden in Great Britain. *British Journal of Cancer,* 107**,** S3–S7. http://dx.doi.org/10.1038/bjc.2012.112

SADHRA, S. S., KURMI, O. P., CHAMBERS, H., LAM, K. B. H. & FISHWICK, D. 2016. Development of an occupational airborne chemical exposure matrix. *Occupational Medicine,* 66**,** 358–364. 10.1093/occmed/kqw027

SAFIZADEH, F., MANDIC, M., SCHöTTKER, B., HOFFMEISTER, M. & BRENNER, H. 2025. Central obesity may account for most of the colorectal cancer risk linked to obesity: evidence from the UK Biobank prospective cohort. *International Journal of Obesity,* 49**,** 619–626. https://doi.org/10.1038/s41366-024-01680-7

SAHA, E., GUEBILA, B. M., FANFANI, V., FISCHER, J., SHUTTA, H. K., MANDROS, P., et al.LOPES-RAMOS, M. C. 2024. Gene regulatory networks reveal sex difference in lung adenocarcinoma. *Biology of Sex Differences,* 15. https://doi.org/10.1186/s13293-024-00634-y

SALARI, N., GHASEMI, H., FATAHIAN, R., MANSOURI, K., DOKANEHEIFARD, S., SHIRI, H. M., et al.MOHAMMADI, M. 2023. The global prevalence of primary central nervous system tumors: a systematic review and meta-analysis. *European Journal of Medical Research,* 28. https://doi.org/10.1186/s40001-023-01011-y

SANIKINI, H., MULLER, C. D., CHADEAU-HYAM, M., MURPHY, N., GUNTER, J. M. & CROSS, J. A. 2020a. Anthropometry, body fat composition and reproductive factors and risk of oesophageal and gastric cancer by subtype and subsite in the UK Biobank cohort. *PLOS ONE,* 15**,** e0240413. https://doi.org/10.1371/journal.pone.0240413

SANIKINI, H., MULLER, C. D., SOPHIEA, M., RINALDI, S., AGUDO, A., DUELL, J. E., et al.CROSS, J. A. 2020b. Anthropometric and reproductive factors and risk of esophageal and gastric cancer by subtype and subsite: Results from the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. *International Journal of Cancer,* 146**,** 929–942. https://doi.org/10.1002/ijc.32386

SCELO, G. & LAROSE, L. T. 2018. Epidemiology and Risk Factors for Kidney Cancer. *Journal of Clinical Oncology,* 36**,** 3574–3581. https://doi.org/10.1200/JCO.2018.79.1905

SCELO, G., LI, P., CHANUDET, E. & MULLER, C. D. 2018. Variability of Sex Disparities in Cancer Incidence over 30 Years: The Striking Case of Kidney Cancer. *European Urology Focus,* 4**,** 586–590. https://doi.org/10.1016/j.euf.2017.01.006

SCHATZKIN, A., SUBAR, F. A., THOMPSON, E. F., HARLAN, C. L., TANGREA, J., HOLLENBECK, R. A., et al.KIPNIS, V. 2001. Design and Serendipity in Establishing a Large Cohort with Wide Dietary Intake Distributions. *American Journal of Epidemiology,* 154**,** 1119–1125. https://doi.org/10.1093/aje/154.12.1119

SCHOENFELD, D. A. 1983. Sample-Size Formula for the Proportional-Hazards Regression Model. *Biometrics,* 39**,** 499. https://doi.org/10.2307/2531021

SCHORR, M., DICHTEL, E. L., GERWECK, V. A., VALERA, D. R., TORRIANI, M., MILLER, K. K. & BREDELLA, A. M. 2018. Sex differences in body composition and association with cardiometabolic risk. *Biology of Sex Differences,* 9. https://doi.org/10.1186/s13293-018-0189-3

SCHUBART, R. J., ELIASSEN, H. A., SCHILLING, A. & GOLDENBERG, D. 2021. Reproductive Factors and Risk of Thyroid Cancer in Women: An Analysis in the Nurses' Health Study II. *Women's Health Issues,* 31**,** 494–502. https://doi.org/10.1016/j.whi.2021.03.008

SERGENTANIS, N. T., ZAGOURI, F., TSILIMIDOS, G., TSAGIANNI, A., TSELIOU, M., DIMOPOULOS, A. M. & PSALTOPOULOU, T. 2015. Risk Factors for Multiple Myeloma: A Systematic Review of Meta-Analyses. *Clinical Lymphoma Myeloma and Leukemia,* 15**,** 563–577.e3. https://doi.org/10.1016/j.clml.2015.06.003

SHAFFER, E. & HUNDAL, R. 2014. Gallbladder cancer: epidemiology and outcome. *Clinical Epidemiology***,** 99. https://doi.org/10.2147/clep.s37357

SHELTON, J., ZOTOW, E., SMITH, L., JOHNSON, S. A., THOMSON, C. S., AHMAD, A., et al.FORMAN, D. 2024. 25 year trends in cancer incidence and mortality among adults aged 35-69 years in the UK, 1993-2018: retrospective secondary analysis. *BMJ,* 384**,** e076962. https://doi.org/10.1136/bmj-2023-076962

SHIPTON, D., TAPPIN, M. D., VADIVELOO, T., CROSSLEY, A. J., AITKEN, A. D. & CHALMERS, J. 2009. Reliability of self reported smoking status by pregnant women for estimating smoking prevalence: a retrospective, cross sectional study. *BMJ,* 339**,** b4347–b4347. https://doi.org/10.1136/bmj.b4347

SHOLL, L. M. 2016. The Molecular Pathology of Lung Cancer. *Surgical Pathology Clinics,* 9**,** 353–378. https://doi.org/10.1016/j.path.2016.04.003

SMEDBY, K. E., MELBYE, M. & ADAMI, H. O. 2018. Non-Hodgkin Lymphoma. *In:* ADAMI, H.-O., HUNTER, D. J., LAGIOU, P. & MUCCI, L. (eds.) *Textbook of Cancer Epidemiology.* 3 ed. New York: Oxford University Press.

SMITH, P. H., BESSETTE, A. J., WEINBERGER, A. H., SHEFFER, C. E. & MCKEE, S. A. 2016. Sex/gender differences in smoking cessation: A review. *Preventive Medicine,* 92**,** 135–140. https://doi.org/10.1016/j.ypmed.2016.07.013

SMYTH, C. E., LAGERGREN, J., FITZGERALD, C. R., LORDICK, F., SHAH, A. M., LAGERGREN, P. & CUNNINGHAM, D. 2017. Oesophageal cancer. *Nature Reviews Disease Primers,* 3. https://doi.org/10.1038/nrdp.2017.48

SONG, M.-A., BENOWITZ, N. L., BERMAN, M., BRASKY, T. M., CUMMINGS, K. M., HATSUKAMI, D. K., et al.SHIELDS, P. G. 2017. Cigarette Filter Ventilation and its Relationship to Increasing Rates of Lung Adenocarcinoma. *JNCI: Journal of the National Cancer Institute,* 109. 10.1093/jnci/djx075

SONG, Y., LIU, X., CHENG, W., LI, H. & ZHANG, D. 2022. The global, regional and national burden of stomach cancer and its attributable risk factors from 1990 to 2019. *Scientific Reports,* 12. https://doi.org/10.1038/s41598-022-15839-7

STATACORP 2025. Stata Statistical Software: Release 19. College Station, TX: StataCorp LLC.

STEEG, V. G. L. & KLEIN, L. S. 2016. SeXX Matters in Infectious Disease Pathogenesis. *PLOS Pathogens,* 12**,** e1005374. https://doi.org/10.1371/journal.ppat.1005374

STIEBAHL, S. 2024. Alcohol statistics: England.Available: https://researchbriefings.files.parliament.uk/documents/CBP-7626/CBP-7626.pdf.

SUDLOW, C., GALLACHER, J., ALLEN, N., BERAL, V., BURTON, P., DANESH, J., et al.COLLINS, R. 2015. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine,* 12**,** e1001779. https://doi.org/10.1371/journal.pmed.1001779

SUN, L., WANG, Z., LU, T., MANOLIO, T. A. & PATERSON, A. D. 2023. eXclusionarY: 10 years later, where are the sex chromosomes in GWASs? *The American Journal of Human Genetics,* 110**,** 903–912. 10.1016/j.ajhg.2023.04.009

SUNG, H., FERLAY, J., SIEGEL, L. R., LAVERSANNE, M., SOERJOMATARAM, I., JEMAL, A. & BRAY, F. 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians,* 71**,** 209–249. https://doi.org/10.3322/caac.21660

SWANTON, C., HILL, W., LIM, E., LEE, C., WEEDEN, C. E., AUGUSTINE, M., et al.JAMAL-HANJANI, M. 2022. LBA1 Mechanism of action and an actionable inflammatory axis for air pollution induced non-small cell lung cancer: Towards molecular cancer prevention. *Annals of Oncology,* 33**,** S1413. https://doi.org/10.1016/j.annonc.2022.08.046

TAUBES, G. 1995. Epidemiology Faces Its Limits. *Science,* 269**,** 164–169. https://doi.org/10.1126/science.7618077

THE ALL OF US RESEARCH PROGRAM INVESTIGATORS 2019. The “All of Us” Research Program. *New England Journal of Medicine,* 381**,** 668–676. https://doi.org/10.1056/NEJMsr1809937

THE TOBACCO AND GENETICS CONSORTIUM 2010. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics,* 42**,** 441–447. https://doi.org/10.1038/ng.571

THE WOMEN'S HEALTH INITIATIVE 1998. Design of the Women's Health Initiative Clinical Trial and Observational Study. *Controlled Clinical Trials,* 19**,** 61–109. https://doi.org/10.1016/S0197-2456(97)00078-0

THEODORESCU, D., LI, Z. & LI, X. 2022. Sex differences in bladder cancer: emerging data and call to action. *Nature Reviews Urology,* 19**,** 447–449. https://doi.org/10.1038/s41585-022-00591-4

THIéBAUT, M. C. A. & BéNICHOU, J. 2004. Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. *Statistics in Medicine,* 23**,** 3803–3820. https://doi.org/10.1002/sim.2098

THOMPSON, D. J., GENOVESE, G., HALVARDSON, J., ULIRSCH, J. C., WRIGHT, D. J., TERAO, C., et al.PERRY, J. R. B. 2019. Genetic predisposition to mosaic Y chromosome loss in blood. *Nature,* 575**,** 652–657. https://doi.org/10.1038/s41586-019-1765-3

THUN, J. M., CARTER, D. B., FESKANICH, D., FREEDMAN, D. N., PRENTICE, R., LOPEZ, D. A., et al.GAPSTUR, M. S. 2013. 50-Year Trends in Smoking-Related Mortality in the United States. *New England Journal of Medicine,* 368**,** 351–364. https://doi.org/10.1056/NEJMsa1211127

THUN, M., HENLEY, J. & TRAVIS, W. D. 2017. Lung Cancer. *In:* THUN, M., LINET, M. S., CERHAN, J. R., HAIMAN, C. A. & SCHOTTENFELD, D. (eds.) *Cancer Epidemiology and Prevention (4th edn)*. Oxford University Press.

TIAN, J., ZUO, C., LIU, G., CHE, P., LI, G., LI, X. & CHEN, H. 2020. Cumulative evidence for the relationship between body mass index and the risk of esophageal cancer: An updated meta-analysis with evidence from 25 observational studies. *J Gastroenterol Hepatol,* 35**,** 730–743. https://doi.org/10.1111/jgh.14917

TORRE, A. L., SIEGEL, L. R., ISLAMI, F., BRAY, F. & JEMAL, A. 2018. Worldwide Burden of and Trends in Mortality From Gallbladder and Other Biliary Tract Cancers. *Clinical Gastroenterology and Hepatology,* 16**,** 427–437. https://doi.org/10.1016/j.cgh.2017.08.017

TOSAKOON, S., LAWRENCE, R. W., SHIELS, S. M. & JACKSON, S. S. 2024. Sex Differences in Cancer Incidence Rates by Race and Ethnicity: Results from the Surveillance, Epidemiology, and End Results (SEER) Registry (2000–2019). *Cancers,* 16**,** 989. https://doi.org/10.3390/cancers16050989

TOWNSEND, P., PHILLIMORE, P. & BEATTIE, A. 1988. *Health and Deprivation: Inequality and the North,* London, Routledge.

TRAN, Q.-L. & DAVIES, L. 2023. Thyroid cancer incidence differences between men and women. *Current Opinion in Endocrine and Metabolic Research,* 31**,** 100472. https://doi.org/10.1016/j.coemr.2023.100472

TRANSACTUAL UK 2021. Trans Lives Survey 2021: Enduring the UK's hostile environment. United Kingdom.Available: https://transactual.org.uk/trans-lives-21/.

TRAVIS, R. C., BALKWILL, A., FENSOM, G. K., APPLEBY, P. N., REEVES, G. K., WANG, X.-S., et al.BERAL, V. 2016. Night Shift Work and Breast Cancer Incidence: Three Prospective

Studies and Meta-analysis of Published Studies. *Journal of the National Cancer Institute,* 108**,** djw169. 10.1093/jnci/djw169

TUDOR, K., TEARNE, S., JEBB, S. A., LEWIS, A., ADAB, P., BEGH, R., et al.AVEYARD, P. 2021. Gender differences in response to an opportunistic brief intervention for obesity in primary care: Data from the BWeL trial. *Clinical Obesity,* 11. 10.1111/cob.12418

U.S. GOVERNMENT ACCOUNTABILITY OFFICE 2001. Drug Safety: Most Drugs Withdrawn in Recent Years Had Greater Health Risks for Women [Internet]. Washington (DC).Available: https://www.gao.gov/products/gao-01-286r.

UK BIOBANK. 2007. *UK Biobank Ethics and Governance Framework* [Online]. Available: https://www.ukbiobank.ac.uk/wp-content/uploads/2025/01/Ethics-and-governance-framework.pdf.

UK BIOBANK. 2011. *UK Biobank: Body Composition Measurement (Resource 1421)* [Online]. UK Biobank. Available: https://biobank.ndph.ox.ac.uk/ukb/refer.cgi?id=1421.

UK BIOBANK. 2014. *UK Biobank: Imaging [Category 100003]* [Online]. UK Biobank. Available: https://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=100003.

UK BIOBANK 2017. UK Biobank: Showcase [Online]. Available: https://biobank.ndph.ox.ac.uk/showcase/.

UK BIOBANK. 2023a. *UK Biobank Cancer Data: linkage from national cancer registries*. [Online]. Available: https://biobank.ctsu.ox.ac.uk/crystal/ukb/docs/CancerLinkage.pdf.

UK BIOBANK. 2023b. *UK Biobank: data providers and dates of data availability.* [Online]. Available: https://biobank.ndph.ox.ac.uk/ukb/exinfo.cgi?src=Data_providers_and_dates.

UK BIOBANK COORDINATING CENTRE. 2006. *UK Biobank: Protocol for a large-scale prospective epidemiological resource* [Online]. Cheshire, UK. Available: https://www.ukbiobank.ac.uk/wp-content/uploads/2025/01/Main-study-protocol.pdf.

VECCHIA, L. C., BOSETTI, C. & ADAMI, H. O. 2018. Thyroid Cancer. *In:* ADAMI, H.-O., HUNTER, D. J., LAGIOU, P. & MUCCI, L. (eds.) *Textbook of Cancer Epidemiology.* 3 ed. New York: Oxford University Press.

VINEIS, P. & SIMONATO, L. 1991. Proportion of Lung and Bladder Cancers in Males Resulting from Occupation: A Systematic Approach. *Archives of Environmental Health: An International Journal,* 46**,** 6–15. https://doi.org/10.1080/00039896.1991.9937423

WALTER, B. R., BRASKY, M. T., BUCKLEY, A. S., POTTER, D. J. & WHITE, E. 2013. Height as an Explanatory Factor for Sex Differences in Human Cancer. *JNCI: Journal of the National Cancer Institute,* 105**,** 860–868. https://doi.org/10.1093/jnci/djt102

WARD, C. L. 2019. Bioelectrical impedance analysis for body composition assessment: reflections on accuracy, clinical utility, and standardisation. *European Journal of Clinical Nutrition,* 73**,** 194–199. https://doi.org/10.1038/s41430-018-0335-3

WATLING, Z. C., KELLY, K. R., WATTS, L. E., GRAUBARD, I. B., PETRICK, L. J., MATTHEWS, E. C. & MCGLYNN, A. K. 2025. Total testosterone, sex hormone-binding globulin, and free testosterone concentrations and risk of primary liver cancer: A prospective analysis of 200,000 men and 180,000 postmenopausal women. *International Journal of Cancer,* 156**,** 1518–1528. https://doi.org/10.1002/ijc.35244

WATTS, L. E., PEREZ-CORNAGO, A., KNUPPEL, A., TSILIDIS, K. K., KEY, J. T. & TRAVIS, C. R. 2021. Prospective analyses of testosterone and sex hormone-binding globulin with the risk of 19 types of cancer in men and postmenopausal women in UK Biobank. *International Journal of Cancer,* 149**,** 573–584. https://doi.org/10.1002/ijc.33555

WHITE, A., IRONMONGER, L., STEELE, C. J. R., ORMISTON-SMITH, N., CRAWFORD, C. & SEIMS, A. 2018. A review of sex-related differences in colorectal cancer incidence, screening uptake, routes to diagnosis, cancer stage and survival in the UK. *BMC Cancer,* 18. https://doi.org/10.1186/s12885-018-4786-7

WILKINSON, R. J., MORRIS, A. J. E., DOWNING, A., FINAN, J. P., ARAVANI, A., THOMAS, D. J. & SEBAG-MONTEFIORE, D. 2014. The rising incidence of anal cancer in <scp>E</scp>ngland 1990–2010: a population-based study. *Colorectal Disease,* 16**,** O234–O239. https://doi.org/10.1111/codi.12553

WILLIAMS, A. L., RICHARDSON, M., MARCOTTE, L. E., POYNTER, N. J. & SPECTOR, G. L. 2019. Sex ratio among childhood cancers by single year of age. *Pediatric Blood Cancer,* 66**,** e27620. https://doi.org/10.1002/pbc.27620

WILUNDA, C., SAWADA, N., YAMAJI, T., IWASAKI, M., INOUE, M. & TSUGANE, S. 2021. Reproductive Factors and Lung Cancer Risk among Never-Smoking Japanese Women with 21 Years of Follow-Up: A Cohort Study. *Cancer Epidemiology, Biomarkers & Prevention,* 30**,** 1185–1192. https://doi.org/10.1158/1055-9965.EPI-20-1399

WISE, A. L., GYI, L. & MANOLIO, T. A. 2013. eXclusion: Toward Integrating the X Chromosome in Genome-wide Association Analyses. *The American Journal of Human Genetics,* 92**,** 643–647. 10.1016/j.ajhg.2013.03.017

WONG, C. S. M., HUANG, J., LOK, V., WANG, J., FUNG, F., DING, H. & ZHENG, Z.-J. 2021. Differences in Incidence and Mortality Trends of Colorectal Cancer Worldwide Based on Sex, Age, and Anatomic Location. *Clinical Gastroenterology and Hepatology,* 19**,** 955–966.e61. https://doi.org/10.1016/j.cgh.2020.02.026

WONG, S. C. M., JIANG, Y. J., GOGGINS, B. W., LIANG, M., FANG, Y., FUNG, H. D. F., et al.CHAN, Y. L. H. 2017. International incidence and mortality trends of liver cancer: a global profile. *Scientific Reports,* 7**,** 45846. https://doi.org/10.1038/srep45846

WONG, Y. Y. J., BASSIG, A. B., LOFTFIELD, E., HU, W., FREEDMAN, D. N., JI, B.-T., et al.LAN, Q. 2020. White Blood Cell Count and Risk of Incident Lung Cancer in the UK Biobank. *JNCI Cancer Spectrum,* 4. https://doi.org/10.1093/jncics/pkz102

WOODWARD, M., TUNSTALL-PEDOE, H., SMITH, W. C. S. & TAVENDALE, R. 1991. Smoking characteristics and inhalation biochemistry in the Scottish population. *Journal of Clinical Epidemiology,* 44**,** 1405–1410. https://doi.org/10.1016/0895-4356(91)90101-e

WORLD CANCER RESEARCH FUND 2018. Body fatness and weight gain and the risk of cancer.Available: https://www.wcrf.org/wp-content/uploads/2024/10/Body-fatness-and-weight-gain_0.pdf.

WORLD HEALTH ORGANIZATION 2013. *International Classification of diseases for Oncology, 3rd Edition: First Revision,* Geneva, World Health Organization.

WORLD HEALTH ORGANIZATION. 2019. *International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10)* [Online]. Geneva: World Health Organization. Available: https://icd.who.int/browse10/2019/en.

WORLD HEALTH ORGANIZATION 2021. WHO global report on trends in prevalence of tobacco use 2000–2025. Geneva.Available: https://www.who.int/publications/i/item/9789240039322.

WORLD HEALTH ORGANIZATION. 2024. *Alcohol* [Online]. Geneva: World Health Organization. Available: https://www.who.int/news-room/fact-sheets/detail/alcohol.

WORLD HEALTH ORGANIZATION. 2026. *Cancer* [Online]. Lyon, France: WHO 2026. Available: https://www.who.int/health-topics/cancer#tab=tab_1.

XIE, S.-H., NESS-JENSEN, E., RABBANI, S., LANGSETH, H., GISLEFOSS, E. R., MATTSSON, F. & LAGERGREN, J. 2020. Circulating Sex Hormone Levels and Risk of Esophageal Adenocarcinoma in a Prospective Study in Men. *American Journal of Gastroenterology,* 115**,** 216–223. https://doi.org/10.14309/ajg.0000000000000446

YALCINKAYA, A., YALCINKAYA, R., SARDH, F. & LANDEGREN, N. 2025. Immune dynamics throughout life in relation to sex hormones and perspectives gained from gender-affirming hormone therapy. *Frontiers in Immunology,* 15. https://doi.org/10.3389/fimmu.2024.1501364

YE, W., NYREN, O. & ADAMI, H. O. 2018. Stomach Cancer. *In:* ADAMI, H.-O., HUNTER, D. J., LAGIOU, P. & MUCCI, L. (eds.) *Textbook of Cancer Epidemiology.* 3 ed. New York: Oxford University Press.

YUAN, Y., LIU, L., CHEN, H., WANG, Y., XU, Y., MAO, H., et al.LIANG, H. 2016. Comprehensive Characterization of Molecular Differences in Cancer between Male and Female Patients. *Cancer Cell,* 29**,** 711–722. https://doi.org/10.1016/j.ccell.2016.04.001

ZHANG, R. E., PFEIFFER, M. R., AUSTIN, A., CLARKE, A. M., HAYES, J., HORNER, M.-J., et al.SHIELS, S. M. 2022. Impact of HIV on Anal Squamous Cell Carcinoma Rates in the United States, 2001-2015. *JNCI: Journal of the National Cancer Institute*. https://doi.org/10.1093/jnci/djac103

ZHANG, T., JOUBERT, P., ANSARI-POUR, N., ZHAO, W., HOANG, H. P., LOKANGA, R., et al.LANDI, T. M. 2021. Genomic and evolutionary classification of lung cancer in never smokers. *Nature Genetics,* 53**,** 1348–1359. https://doi.org/10.1038/s41588-021-00920-0

ZHANG, X., CHO, E. & ADAMI, H. O. 2018. Kidney Cancer. *In:* ADAMI, H.-O., HUNTER, D. J., LAGIOU, P. & MUCCI, L. (eds.) *Textbook of Cancer Epidemiology.* 3 ed. New York: Oxford University Press.

ZHAO, J., WANG, Q., TAN, F. A., LOH, L. J. C. & TOH, C. H. 2024. Sex differences in cancer and immunotherapy outcomes: the role of androgen receptor. *Frontiers in Immunology,* 15. https://doi.org/10.3389/fimmu.2024.1416941

ZHAO, J., XU, L., SUN, J., SONG, M., WANG, L., YUAN, S., et al.LI, X. 2023. Global trends in incidence, death, burden and risk factors of early-onset cancer from 1990 to 2019. *BMJ Oncology,* 2**,** e000049. https://doi.org/10.1136/bmjonc-2023-000049

ZHENG, D., TRYNDA, J., WILLIAMS, C., VOLD, J. A., NGUYEN, J. H., HARNOIS, D. M., et al.LI, Z. 2019. Sexual dimorphism in the incidence of human cancers. *BMC Cancer,* 19**,** 684. https://doi.org/10.1186/s12885-019-5902-z