Addressing Psychosocial and Lifestyle Risk Factors to Promote Primary Cancer Prevention: an integrated platform to promote behavioural change (IBeCHANGE)

Project Number: 101136840

D3.2 – Analysis of publicly available data

Related Work Package	**WP3 – Novel approaches for interaction through a Virtual User Model**
Related Task	**T3.2 – Collection and analysis of publicly available survey and self-reported data**
Lead Beneficiary	SD
Contributing Beneficiaries	IEO, UNIPA
Document version	v1.0
Deliverable type	R – Document, report
Dissemination level	PU - Public
Due date	30/11/2024
Delivery date	DD/MM/YEAR

Authors	Aline Machiavelli (SD), Ana Beatriz Lopes (SD), Mirtha Valencia (SD), Antonio Lindenberg (SD)
Contributors	Ricardo Pietrobon (SD), Anna García Serra (ICO), Gabriella Pravettoni, Marianna Masiero, Elisa Tomezzoli, Giorgia Miale (IEO)
Reviewers	iBeChange Consortium

Funded by
the European Union

This project has received funding from the European Union’s Horizon Europe research and innovation programme under the Grant Agreement Number 101136840.

Disclaimer

Table of Contents

List of Abbreviations

Abbreviation	Explanation
BCAC	Breast Cancer Association Consortium
BMI	Body Mass Index
CDM	Common Data Model
CAPI	Computer-Assisted Personal Interviewing
CI5	Cancer Incidence in Five Continents
COSM	Cohort of Swedish Men
DICA	Dutch Institute for Clinical Auditing
DLCA	Dutch Lung Cancer Audit
DLSA	Dutch Lung Surgery Audit
DSCA	Dutch Surgical Colorectal Audit
ECIS	European Cancer Information System
EHIS	European Health Interview Survey
EUROCARE	European Cancer Registry-based Study on Survival and Care
EU-SILC	European Union statistics on income and living conditions
GCO	Global Cancer Observatory
GDC	Genomic Data Commons
HFA-DB	European Health for All database
IHS	Irish Health Survey
INE	Instituto Nacional de Estadística (National Statistics Institute)
ISSDA	Irish Social Science Data Archive
NCR	National Cancer Registry
NORDCAN	Nordic Cancer Database
NVI	Nacionalinio vėžio instituto (Lithuanian Cancer Registry)
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
QoL	Quality of Life

SLORA	SLOvenski RAk Register (Slovenia and Cancer Registry)
SMD	Standardised Mean Difference
SND	Swedish National Data Service
WHO	World Health Organization
WLH	Women’s Lifestyle and Health
ZfKD	Zentrum für Krebsregisterdaten (German Centre for Cancer Registry Data)

List of Tables

List of Figures

Executive Summary

This report presents the results of the work conducted under Task 3.2 of the iBeChange project, with the goal of collecting and analysing publicly available data on cancer risk factors. Focused on understanding how behavioural and psychosocial risk factors impact health and cancer outcomes, such as incidence and survivorship, the task involved identifying oncology-related public data sources across EU member states. These data sources included cancer registries, surveys, and other quantitative sources. These datasets are instrumental in exploring connections between lifestyle factors (such as diet and smoking) and chronic disease risks or ageing outcomes.

Following a screening of 3,505 articles, we identified 59 eligible databases and secured access to 28 of them. This report includes descriptions of these 28 databases, with key variables standardised using OHDSI's OMOP Common Data Model (CDM). From this initial pool, five datasets were selected for in-depth analysis based on their relevance and data quality, enabling us to uncover discrepancies in the reporting of health and cancer risk factors across EU member states.

Key findings reveal significant links between lifestyle choices and cancer outcomes, with variables like age, marital status, and smoking patterns emerging as strong predictors. Aggregate data across member countries also highlighted differences in diet, exercise, and protective behaviours like sun protection, underscoring the diversity of lifestyle influences on cancer prevalence. These findings support the iBeChange goal to develop a personalised prevention tool and provide data-driven insights to inform primary cancer prevention.

1. Introduction

Work Package 3 (WP3) of the iBeChange project focuses on developing novel approaches for interaction through a Virtual User Model. The aim is to implement personalised and data-driven strategies that maximise user acceptance and adherence by collecting and analysing retrospective and publicly available data related to behavioural and psychosocial risk factors. This information will inform the iBeChange Platform (WP4), contributing to the development of personalised interventions and interfaces to enhance user experience.

Deliverable 3.2 presents the results of the analysis of publicly available surveys and self-reported data under Task 3.2. This task involves the identification and examination of publicly available oncology datasets across member countries, including cancer registries, patient feedback, and other qualitative data sources. We explore key variables like tobacco use, diet, psychosocial determinants of health, and overall quality of life (QoL). This analysis aims to clarify the relationship between lifestyle factors and cancer risk, providing essential findings for the personalised interventions that will be incorporated into the iBeChange platform.

In this deliverable, we document the methodology, datasets, and findings from the analysis. In line with WP3’s goal, we aim to advance cancer prevention and survivorship by developing data-driven, personalised health interventions. By fostering collaboration and knowledge exchange with related projects, we strengthen our efforts to improve outcomes for cancer patients through innovative, user-centred solutions based on comprehensive data analysis.

2. Methodology: data mining and standardisation process

This chapter outlines the methodology of a comprehensive data mining and standardisation process for task T3.2, “Collection and analysis of publicly available survey and self-reported data” for the iBeChange project. Our primary goal was to identify, access, and standardise public databases that contain information related to cancer (breast, colorectal, and lung) and key behavioural and psychosocial risk factors such as smoking and education attainment levels within European Union countries. This part of the project involved multiple stages, from initial literature reviews to data extraction, cleaning, and transformation across various datasets.

2.1 Searching publicly available databases

In order to conduct an efficient search of publicly available databases, we developed a function in R (R version 4.4.1 from 2024-06-14) utilising the RISmed package.[1] This function, named "pubmedDatabaseSearch," enabled us to locate a variety of studies and extract relevant information such as titles, authors, affiliations, dates, and abstract snippets based on a specified query. The source code for the function is available in Appendix 1.

The function is composed of the following arguments:

query: The search query.

base_terms: A set of key terms used to identify specific databases (default values include: registry, database, dataset, claim, cohort, and study).

remove_terms: A set of terms to exclude from the search (default: NULL).

countries: A set of countries to include in the search (default: NULL).

include_base_terms: A logical flag indicating whether to include base_terms in the search (default: TRUE).

nmax: The maximum number of records to retrieve (default: 1000).

nwords: The number of words surrounding the keywords in the snippets (default: 10).

filename: The name of the file to save the results. If set to NULL, a filename will be generated automatically (default: NULL).

silent: A logical flag to suppress messages during function execution (default: FALSE).

Our search was guided by a combination of three key factors: outcomes, cancer types, and geographic constraints. For outcomes, we focused on "quality of life," "diagnosis," and "survival," which are essential for understanding patient trajectories. The targeted cancer types included breast, colorectal, and lung cancer, as well as specific subtypes like colon and rectal cancer. To align with the scope of the iBeChange project, we limited our search to articles from European countries, specifying a list of nations within the European Union.

Additionally, we refined the search to include articles that mention data sources. We used terms such as "database," "dataset," "registry," "repository," and "consortium" to ensure we captured publications likely to reference public databases. We also excluded systematic reviews, meta-

analyses, and other secondary studies using the “remove_terms” parameter, as these articles typically do not provide direct access to primary datasets. Finally, the function call generates a CSV file containing the results, with a maximum limit of 99,999 articles, to ensure a comprehensive search. The following R code outlines the query parameters and constraints for collecting the articles.

```
outcomes <- '("quality of life" OR diagnosis OR survivors)'
cancer_type <- '("breast cancer" OR "colorectal cancer" OR "lung cancer" OR "colon cancer" OR "rectum cancer")'
country_list <- '("Austria" OR "Belgium" OR "Bulgaria" OR "Croatia" OR "Cyprus" OR "Czech Republic" OR "Denmark" OR "Estonia" OR "Finland" OR "France" OR "Germany" OR "Greece" OR "Hungary" OR "Ireland" OR "Italy" OR "Latvia" OR "Lithuania" OR "Luxembourg" OR "Malta" OR "Netherlands" OR "Poland" OR "Portugal" OR "Romania" OR "Slovakia" OR "Slovenia" OR "Spain" OR "Sweden")'


query <- paste(outcomes, "AND", cancer_type, "AND", country_list)
base_termers <- c("database", "dataset", "registry", "repository", "consortium")
remove_terms <- c("systematic review", "meta-analysis", "meta-review", "Medline", "Embase", "Cochrane", "Ovid", "Scopus")
countries <- c(" "," ")
format_query <- "Teste query3"
filename <- paste0("data/pubMed/", format_query, ".csv")


pubmedDatabaseSearch(
    query,
    base_terms = base_termers,
    filename = filename,
    remove_terms = remove_terms,
    countries = NULL,
    include_base_terms = TRUE,
    nmax = 99999
)
```

We also performed a manual search using ChatGPT and Google. However, most databases found during this search matched those already mentioned in the papers identified using the R function.

2.2 Article and database selection

After obtaining the initial list of articles, we carefully screened these studies to identify relevant databases. We focused on databases related to colorectal, breast, and lung cancer that also included lifestyle risk factors or psychosocial variables such as alcohol use, tobacco use, physical activity, diet, depression, and anxiety, among others (see the complete list in Table 1).

Table 1: List of lifestyle and psychosocial risk factors of interest.

Risk Factors	
Socio-demographic characteristics	Age
	Sex/gender
	Education
	Marital status
	Race/ethnicity
	Socio-economic status (income/financial insecurity)
	Healthcare system factors
	Household (e.g., cohabitation)
	Employment status
Psychiatric and Psychological symptoms	Depression
	Emotional distress
	Anxiety
	Post-traumatic stress disorder
	Alexithymia
Personality Traits	Type A personality
	Type C personality
	Type D personality
	Agreeableness
	Optimism
	Dispositional mindfulness
	Extraversion
	Neuroticism
	Novelty-seeking
	Hostility

	Hardiness
	Grit
	Sense of coherence
Emotional aspects	Affect (positive/negative)
	Loneliness
	Hopelessness
Social aspects	Social attachment
	Social isolation
	Social network
	Social support (Perceived social support; social support satisfaction)
Disease-related aspects	Illness representation (negative illness perception)
	Cognitive symptoms/impairments
	Fatigue
	The mental burden of the disease
	Symptom burden
	Fear of recurrence
Stress-related aspects	Post-traumatic growth
	Stress (perceived stress)
	Traumatic (life) events (cancer diagnosis/having cancer, incidents)
	Stressful (life) events (e.g., divorce, recent loss events, cancer diagnosis/having cancer)
Cognitive aspects	Health Locus of control (HLOC)
	Rumination
	Worry (e.g., financial worries, subjective threat of cancer)
	Self-efficacy (confidence)
	Self-esteem
	Perceived ability to cope
	Future appraisal and perspective
Coping strategies	Problem-solving
	Expressing cancer-related emotions

	Proactive coping
	Active coping strategies (acceptance and reframing)
	Functional coping strategies (combativeness and seeking solutions to problems)
	Problem-centred coping strategies
	Emotion-centered coping strategies
Other psychosocial variables	Resilience
	Body image
	Compliance
	Adherence
	Spirituality
	QoL (e.g., global health status)

2.3 Database access and standardisation

After compiling a curated list of datasets that included our variables of interest and met our inclusion criteria described above, we proceeded with accessing the selected databases. This process required various methods depending on the database. Some were available for direct download from official websites, while others required a formal access request, ranging from an email inquiry to a complete proposal submission. At this stage, we opted to defer access to databases that required a fee; however, if deemed essential for the iBeChange project, we may seek approval from consortium members to pursue them in the future.

The next phase of the project involved standardising the data obtained from the accessed databases. Each of these databases contained multiple datasets, each with varying structures and levels of detail. We catalogued these datasets using Google Sheets, adding dedicated columns to organise the variables present in each dataset (represented in the rows). To ensure a deeper understanding of each dataset, we included detailed metadata to provide essential context, such as risk factors involved, detailed regions, and timeframes. This information is critical for evaluating the quality and relevance of the data, facilitating informed decision-making during analysis.

A critical point of this standardisation process was ensuring variable compatibility across datasets. When the same variable was present in different databases, we harmonised the variable names to create a unified structure that allowed cross-database analysis. To further enhance this, we added the OHDSI's OMOP CDM (Observational Medical Outcomes Partnership Common Data Model)[2,3] concept ID to standardise each variable, when applicable. The Athena[4] tool was employed to search for the OMOP concept IDs. The OMOP concept IDs are particularly valuable as they provide a standardised vocabulary and format that aligns disparate data sources, enabling researchers and healthcare professionals to make more accurate comparisons and analyses across datasets.

Additionally, we structured the database documentation spreadsheet to include a dedicated "Data Overview" tab, with guides to understanding the document, and a "Data Dictionary" tab, detailing each variable's name, description, examples, and OMOP concept ID (if applicable). These additions improve usability to ensure stakeholders can easily navigate and understand the datasets.

2.4 Statistical analyses

We selected five datasets for more detailed analyses based on the type of information they provide, prioritising individual-level data over aggregate data. We chose datasets that included multiple variables of interest, as outlined in Table 1. For each selected dataset, we assessed the frequency and percentage of categorical variables (such as cancer diagnosis, mortality, and gender) and examined the near-zero variance. Additionally, we analysed the distribution of numeric variables, including mortality and cancer rates for aggregate data, along with their associated patterns of missing values.[5] Near-zero variance occurs when a categorical variable presents a low frequency of unique values over the sample size, i.e., the variable is almost constant, and we addressed it by combining different variable categorisations. We conducted comparisons for the exploratory analysis through a standardised mean difference (SMD), i.e., the difference in means or proportions divided by the pooled standard deviation. We will consider the following guidelines when interpreting SMD magnitude: SMD = 0.2 corresponds to a small effect; SMD = 0.5 corresponds to a medium effect; and SMD = 0.8 corresponds to a large effect.[6] We also present *p*-values for *t*-tests (for numeric variables) and Chi-square tests (for categorical variables). We considered *p*-values below 0.05 as statistically significant.

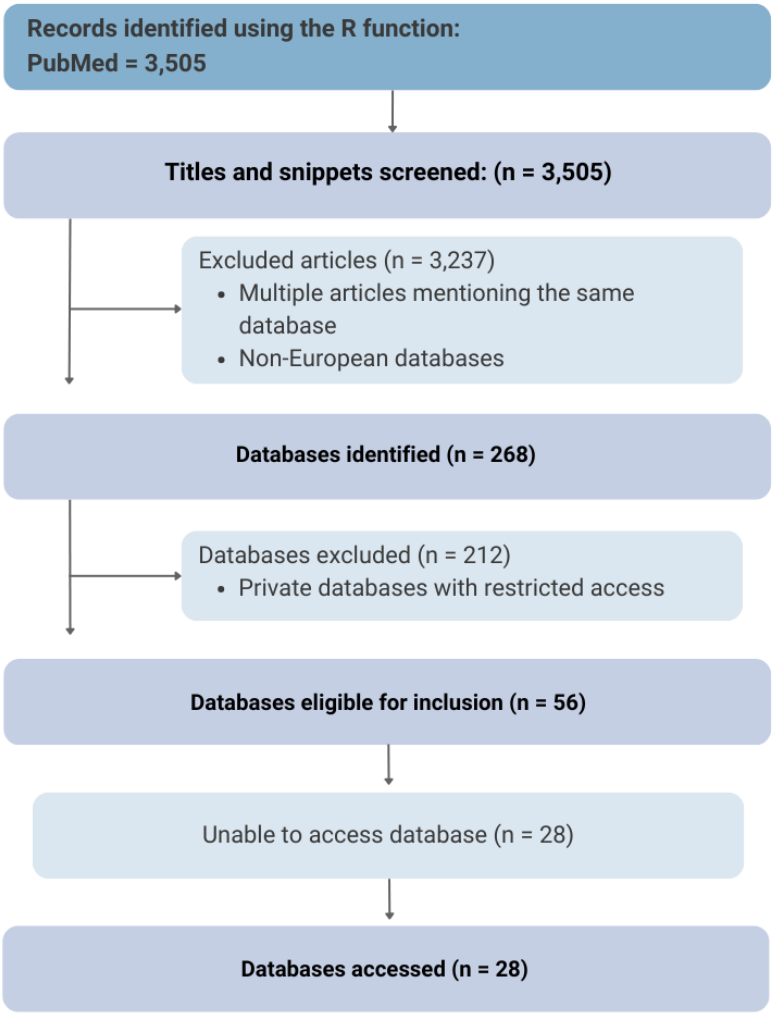
Bivariate choropleth maps were created to classify each region or country based on cancer diagnosis or mortality rates (x-axis) and various risk factors of interest, such as Body Mass Index (BMI), alcohol and tobacco consumption, nutrition, and physical activity (y-axis). We employed a quantile-based classification system with three dimensions.

The association between cancer diagnosis and each risk factor was evaluated using a series of generalised linear models with a binomial family, i.e., logistic regression models. Results are presented as odds ratios for categorical outcomes (cancer diagnosis), accompanied by 95% confidence intervals and *p*-values. We interpreted the results as significant when the confidence interval did not include the value of 1 and the *p*-value was lower than 0.05.

3. Results: databases identified

Our article selection process started with title and abstract screening of the 3,505 articles retrieved using our R function (Figure 1). In this first step, we excluded 3,237 studies that mentioned the same database and those that included non-European databases. We then identified 268 databases and further excluded 212 of them because they were private databases with restricted access. The initial 56 databases we evaluated and attempted to access are described in Appendix 2. Of the 56 eligible databases, we were able to access 28 databases, which formed the basis of our subsequent analysis. These databases were selected based on their relevance to our study's objectives and the comprehensive data they offer regarding cancer incidence, risk factors, and psychosocial aspects. Our focus was mainly on those databases that provide patient-level data, as they offer the granularity necessary for detailed analysis to develop targeted cancer prevention strategies. We describe these databases in detail below.

Figure 1. Flowchart illustrating the process of selecting articles and databases.



Out of the 28 downloaded databases, 25 provided aggregate data at the country or regional level. Three offered individual-level data that allowed for in-depth analyses of individual risk factors and outcomes. The datasets with individual-level data included the Healthy Ireland Survey, the Irish Health Survey, and the EU statistics on income and living conditions (EU-SILC).

First, we prepared detailed documentation and standardised all 28 databases to which we obtained access. During this approach, we identified 5,626 datasets containing 426 unique variables. The detailed documentation with the standardisation of these databases is available elsewhere (Link: https://docs.google.com/spreadsheets/d/1F6fumAyuQ6UzO2Fm6hg0dn2uXNSnrOV8M5oYy3w8Gzw/edit?gid=1373898473#gid=1373898473).

After completing the documentation of all 28 databases, we selected five databases to conduct more in-depth analyses. Among the five selected databases, three contain individual-level data, and two contain aggregate data. Individual-level data allows for detailed examination of individual characteristics, variability, and outcomes, enabling more precise analyses of risk factors associated with cancer. Additionally, aggregate data offers broader geographic coverage and detailed information on multiple countries and our key risk factors of interest. The individual-level data we had access to included information only for two countries, Ireland and Austria. Therefore, we incorporated aggregate data to gain an overview of trends and risk factors across multiple European countries, providing a broader context for our findings.

In the following sections, we will first provide a detailed description of the three individual-level datasets we collected. Next, we will give an overview of the remaining 25 databases to which we gained access. After that, we will describe in detail two of these databases that contain aggregate data, chosen for their extensive coverage of European countries and the availability of behavioural risk factors.

3.1 Individual-level data

3.1.1 Healthy Ireland Survey

The Healthy Ireland Survey is an annual interviewer-administered questionnaire of the health and health behaviours of people living in Ireland, commissioned by the Department of Health and carried out by Ipsos, a global market research and consulting company.[7] The purpose of the Healthy Ireland Survey is to provide current and credible data about factors that determine health and well-being in the population aged 15 years and older. The data is used to monitor key trends and policy impacts in areas such as general health, health service utilisation, and key lifestyle factors, such as smoking, alcohol consumption, diet, and long-term health conditions, including cancer.[8] As cancer remains a leading cause of morbidity and mortality in Ireland,[9] the ability to explore the possible correlations between cancer rates and risk factors is essential to act in the prevention of cancer.

Since its inception in 2015, each questionnaire wave involves a sample of approximately 7,500 randomly selected individuals representative of the Irish population.[10] The 2022 edition, in particular, reflects the evolving health landscape post-COVID-19 pandemic, making it an invaluable database.

Table 2 presents the distribution of cancer incidence across sociodemographic variables, such as age, gender, region, marital status, and risk factors (e.g., tobacco and alcohol consumption). A total of 7,455 participants were included in the analysis, with 164 reporting a cancer diagnosis. Age emerged as a significant factor ($p < 0.001$), with older age groups showing higher cancer prevalence. For example, individuals aged 65-74 years presented a 34.8% cancer rate compared to 0.61% in the 15-24 age group. Additionally, marital status was significantly associated with cancer

($p < 0.001$), where 59.1% of the cancer-diagnosed participants were married or in a civil partnership. There were no significant associations between cancer incidence and gender. Moreover, lifetime alcohol consumption did not show statistical significance. Still, alcohol intake in the last 6 months was demonstrated to increase the likelihood of developing cancer ($p <0.001$), especially among those who drank twice or once a week (29.6%). Furthermore, a significant association was observed among malignancies and past consumers of tobacco products ($p < 0,001$), with 56 cancer participants (37.8%) reporting daily smoking. The survey also indicated a significant association between cancer and the willingness to practice more exercise to lose or maintain weight. Specifically, 40.2% of participants diagnosed with cancer were not willing to increase their exercise, compared to 53.9% of those without a cancer diagnosis ($p < 0.001$). Another relevant sociodemographic variable was the education attainment level. Compared to participants without a cancer diagnosis, a higher percentage of participants with a cancer diagnosis had completed primary education (4.51% vs. 14.7%), lower secondary education (10.8% vs. 15.3%), or upper secondary education (23.5% vs. 27%). However, the trend was reversed for higher levels of education, with a smaller percentage of participants with a cancer diagnosis having obtained a bachelor's degree (31.2% vs. 23.3%) or a postgraduate diploma (13.4% vs. 4.29%).

Table 2: Distribution of cancer incidence across sociodemographic variables and risk factors.

Variable [Missing]	Total (7,455)	Cancer or a malignant tumour (164)	No cancer or a malignant tumour (7,291)	*p*-value
Age category [0]				$p < 0.001$ (Chi-square test)
- 15-24 years	614 (8.24%)	1 (0.61%)	613 (8.41%)	
- 25-34 years	1,168 (15.7%)	5 (3.05%)	1,163 (16%)	
- 35-44 years	1,443 (19.4%)	11 (6.71%)	1,432 (19.6%)	
- 45-54 years	1,500 (20.1%)	23 (14%)	1,477 (20.3%)	
- 55-64 years	1,240 (16.6%)	36 (22%)	1,204 (16.5%)	
- 65-74 years	1,003 (13.5%)	57 (34.8%)	946 (13%)	
- 75 years and over	487 (6.53%)	31 (18.9%)	456 (6.25%)	
Gender [0]				$p = 0.239$ (Chi-square test)
- Female	3,517 (47.2%)	67 (40.9%)	3,450 (47.3%)	
- Male	3,932 (52.7%)	97 (59.1%)	3,835 (52.6%)	
- Other	6 (0.08%)	0 (0%)	6 (0.082%)	
Region [0]				$p = 0.833$ (Chi-square test)
- Conn/Ulster	1,172 (15.7%)	30 (18.3%)	1,142 (15.7%)	
- Dublin	2,381 (31.9%)	51 (31.1%)	2,330 (32%)	
- Munster	1,936 (26%)	42 (25.6%)	1,894 (26%)	
- Rest of Leinster	1,966 (26.4%)	41 (25%)	1,925 (26.4%)	
Marital status [0]				$p < 0.001$ (Chi-square test)

Variable [Missing]	Total (7,455)	Cancer or a malignant tumour (164)	No cancer or a malignant tumour (7,291)	*p*-value
- Single, never married, and never in a civil partnership	2,483 (33.3%)	27 (16.5%)	2,456 (33.7%)	
- Married or In a registered same-sex civil partnership	4,075 (54.7%)	97 (59.1%)	3,978 (54.6%)	
- Re-married	64 (0.858%)	5 (3.05%)	59 (0.809%)	
- Separated	241 (3.23%)	8 (4.88%)	233 (3.2%)	
- Divorced	230 (3.09%)	6 (3.66%)	224 (3.07%)	
- Widowed	362 (4.86%)	21 (12.8%)	341 (4.68%)	
Alcohol consumption in lifetime [9]				p = 0.117 (Chi-square test)
- Yes	5,942 (79.7%)	125 (76.2%)	5,817 (79.9%)	
- Never	880 (11.8%)	18 (11%)	862 (11.8%)	
I have only had a few sips of alcohol in my lifetime	624 (8.37%)	21 (12.8%)	603 (8.28%)	
Frequency of alcohol consumption (last 6 months) [1,526]				p < 0.001 (Chi-square test)
- Daily	133 (1.78%)	3 (2.4%)	130 (2.24%)	
- 5-6 times a week	85 (1.14%)	6 (4.8%)	79 (1.36%)	
- 3-4 times a week	561 (7.53%)	11 (8.8%)	550 (9.48%)	
- Twice or once a week	2,368 (31.8%)	37 (29.6%)	2,331 (40.2%)	
- 2-3 times a month	776 (10.4%)	11 (8.8%)	765 (13.2%)	
- Once or less than once a month	1,321 (17.7%)	24 (19.2%)	1,297 (22.3%)	
- I did not drink in the last 6 months, but I drank longer ago.	685 (9.19%)	33 (26.4%)	652 (11.2%)	
In a typical day, how many drinks (last 6 months) [2,237]				p = 0.219 (Chi-square test)
- 1-2	1,741 (23.4%)	41 (44.6%)	1,700 (33.2%)	
- 3-4	1,617 (21.7%)	22 (23.9%)	1,595 (31.1%)	
- 5-6	811 (10.9%)	14 (15.2%)	797 (15.5%)	
- 7-9	461 (6.18%)	7 (7.61%)	454 (8.86%)	
- 10+	588 (7.89%)	8 (8.7%)	580 (11.3%)	
Frequency of consumption of 6 or more standard drinks on one occasion [2,309]				p = 0.2 (Chi-square test)
- More than once a week	502 (6.73%)	8 (8.79%)	494 (9.77%)	
- Once a week	636 (8.53%)	12 (13.2%)	624 (12.3%)	
- 2/3 times a month	387 (5.19%)	5 (5.49%)	382 (7.56%)	
- Once a month	598 (8.02%)	6 (6.59%)	592 (11.7%)	

Variable [Missing]	Total (7,455)	Cancer or a malignant tumour (164)	No cancer or a malignant tumour (7,291)	*p*-value
- Less than once a month	934 (12.5%)	12 (13.2%)	922 (18.2%)	
- Never	2,089 (28%)	48 (52.7%)	2,041 (40.4%)	
Who were you drinking with? (last day you had an alcoholic drink) [2,215]				p = 0.666 (Chi-square test)
- On my own	563 (7.55%)	12 (13%)	551 (10.7%)	
- With my spouse/partner only	2,013 (27%)	39 (42.4%)	1,974 (38.3%)	
- With one other person (not spouse/partner)	592 (7.94%)	9 (9.78%)	583 (11.3%)	
- In a group of people	2,072 (27.8%)	32 (34.8%)	2,040 (39.6%)	
Where were you drinking [2,214]				p = 0.212 (Chi-square test)
- At home (inside or outside)	2,489 (33.4%)	48 (52.2%)	2,441 (47.4%)	
- In someone else’s home (inside or outside)	449 (6.02%)	2 (2.17%)	447 (8.68%)	
- In/outside a pub or in a nightclub	1,445 (19.4%)	22 (23.9%)	1,423 (27.6%)	
- In/outside a restaurant	748 (10%)	17 (18.5%)	731 (14.2%)	
- In a public place outdoors, while on the move, or somewhere else	38 (0.51%)	1 (1.09%)	37 (0.719%)	
- At a sport/music event, sports club/centre, or at a function/wedding	72 (0.966%)	2 (2.17%)	70 (1.36%)	
Where was this drink purchased [4,497]				p = 0.109 (Chi-square test)
- Pub/bar/hotel/restaurant/nightclub/social club/sports club/venue	81 (1.09%)	2 (3.77%)	79 (2.72%)	
- Supermarket or online	2,124 (28.5%)	40 (75.5%)	2,084 (71.7%)	
- Convenience shop, standalone off-license or wholesaler/supplier	664 (8.91%)	7 (13.2%)	657 (22.6%)	
- Duty-free shop, abroad/away/on holiday, received it as a gift or somewhere else	89 (1.19%)	4 (7.55%)	85 (2.93%)	
The highest number of standard drinks on a single occasion (last 6 months) (Mean/SD) [2,579]	7.52 (+-5.72)	6.4 (+-5.17)	7.54 (+-5.73)	p = 0.044 (t-test)
Emotional nervous or psychiatric problems diagnosis [0]	188 (2.52%)	5 (3.05%)	183 (2.51%)	p = 0.854 (Chi-square test)
Smoke tobacco products [2]				p = 0.138 (Chi-square test)
- Yes, daily	885 (11.9%)	13 (7.93%)	872 (12%)	
- Yes, occasionally	250 (3.35%)	3 (1.83%)	247 (3.39%)	
- No	6,318 (84.7%)	148 (90.2%)	6,170 (84.6%)	

Variable [Missing]	Total (7,455)	Cancer or a malignant tumour (164)	No cancer or a malignant tumour (7,291)	*p*-value
Tobacco products (in the past) [1,139]				p < 0.001 (Chi-square test)
- Yes, daily	1,562 (21%)	56 (37.8%)	1,506 (24.4%)	
- Yes, occasionally	506 (6.79%)	16 (10.8%)	490 (7.94%)	
- No	4,248 (57%)	76 (51.4%)	4,172 (67.6%)	
Time since last smoking [5,387]				p = 0.923 (Chi-square test)
- Within the past month (anytime< than 1 month ago)	26 (0.349%)	1 (1.39%)	25 (1.25%)	
- Within the past 3 months (1 month but less than 3 months ago)	35 (0.469%)	1 (1.39%)	34 (1.7%)	
- Within the past 6 months (3 months but less than 6 months ago)	39 (0.523%)	1 (1.39%)	38 (1.9%)	
- Within the past year (6 months but less than 1 year ago)	72 (0.966%)	0 (0%)	72 (3.61%)	
- Within the past 5 years	361 (4.84%)	13 (18.1%)	348 (17.4%)	
- Within the past 10 years (5 years but less than 10 years ago)	296 (3.97%)	10 (13.9%)	286 (14.3%)	
- 10 or more years ago	1,234 (16.6%)	46 (63.9%)	1,188 (59.5%)	
- Don't Know	4 (0.054%)	0 (0%)	4 (0.2%)	
- Refused	1 (0.013%)	0 (0%)	1 (0.05%)	
Number of manufactured cigarettes per day (Mean/SD) [0]	1.16 (+-4.29)	0.811 (+-3.35)	1.17 (+-4.3)	p = 0.185 (t-test)
Number of hand-rolled cigarettes per day (Mean/SD) [0]	0.381 (+-2.61)	0.305 (+-2.33)	0.382 (+-2.61)	p = 0.675 (t-test)
Number of pipes full of tobacco per day (Mean/SD) [0]	0.003 (+-0.151)	0 (+-0)	0.003 (+-0.152)	p = 0.054 (t-test)
Number of cigars per day (Mean/SD) [0]	0.002 (+-0.134)	0 (+-0)	0.002 (+-0.136)	p = 0.12 (t-test)
Tobacco habits [7]				p = 0.673 (Chi-square test)
- I have never tried e-cigarettes	6,263 (84%)	143 (87.2%)	6,120 (84%)	
- I have tried e-cigarettes but do not use them (anymore)	829 (11.1%)	16 (9.76%)	813 (11.2%)	
- I have tried e-cigarettes and still use them daily	202 (2.71%)	3 (1.83%)	199 (2.73%)	
- I have tried e-cigarettes and still use them occasionally	154 (2.07%)	2 (1.22%)	152 (2.09%)	
Stopped smoking for one day or longer - trying to quit smoking (in the past 12 months) [6,148]				p = 0.129 (Chi-square test)
- Yes	611 (8.2%)	14 (73.7%)	597 (46.4%)	
- No	684 (9.18%)	5 (26.3%)	679 (52.7%)	

Variable [Missing]	Total (7,455)	Cancer or a malignant tumour (164)	No cancer or a malignant tumour (7,291)	*p*-value
- Don't Know	9 (0.121%)	0 (0%)	9 (0.699%)	
- Refused	3 (0.04%)	0 (0%)	3 (0.233%)	
Nicotine patches, gum, lozenges, spray - trying to quit smoking [0]	101 (1.35%)	5 (3.05%)	96 (1.32%)	p = 0.12 (Chi-square test)
Varenicline/Champix or Buproprion/Zyban (prescribed medication) - trying to quit smoking [0]	9 (0.121%)	0 (0%)	9 (0.123%)	p = 1 (Chi-square test)
Acupuncture - trying to quit smoking [0]	4 (0.054%)	0 (0%)	4 (0.055%)	p = 1 (Chi-square test)
Smokers telephone Quitline/Helpline - trying to quit smoking [0]	6 (0.08%)	0 (0%)	6 (0.082%)	p = 1 (Chi-square test)
www.quit.ie - trying to quit smoking [0]	9 (0.121%)	0 (0%)	9 (0.123%)	p = 1 (Chi-square test)
www.facebook.com/HSEquit - trying to quit smoking [0]	4 (0.054%)	0 (0%)	4 (0.055%)	p = 1 (Chi-square test)
E-cigarettes - trying to quit smoking [0]	76 (1.02%)	2 (1.22%)	74 (1.01%)	p = 1 (Chi-square test)
Other aid, help, support - trying to quit smoking [0]	7 (0.094%)	0 (0%)	7 (0.096%)	p = 1 (Chi-square test)
No help used attempted to quit 'cold-turkey' - trying to quit smoking [0]	405 (5.43%)	8 (4.88%)	397 (5.45%)	p = 0.887 (Chi-square test)
Hypnotist - trying to quit smoking [0]	5 (0.067%)	0 (0%)	5 (0.069%)	p = 1 (Chi-square test)
Book - trying to quit smoking [0]	3 (0.04%)	0 (0%)	3 (0.041%)	p = 1 (Chi-square test)
Don't Know - trying to quit smoking [0]	11 (0.148%)	1 (0.61%)	10 (0.137%)	p = 0.596 (Chi-square test)
Plans to quit smoking [6,320]				p = 0.05 (Chi-square test)
- Trying to quit	130 (1.74%)	6 (37.5%)	124 (11.1%)	
- Actively planning to quit	203 (2.72%)	2 (12.5%)	201 (18%)	
- Thinking about quitting but not planning to	345 (4.63%)	4 (25%)	341 (30.5%)	
- Not thinking about quitting	441 (5.92%)	4 (25%)	437 (39.1%)	
- Don't Know	15 (0.201%)	0 (0%)	15 (1.34%)	
- Refused	1 (0.013%)	0 (0%)	1 (0.089%)	
Number of snack foods portions (Mean/SD) [3,086]	1.81 (+-1.07)	1.79 (+-0.904)	1.81 (+-1.07)	p = 0.883 (t-test)
Eat fewer calories to lose/maintain weight [0]	2,150 (28.8%)	48 (29.3%)	2,102 (28.8%)	p = 0.972 (Chi-square test)
Eat less fat to lose/maintain weight [0]	1,497 (20.1%)	32 (19.5%)	1,465 (20.1%)	p = 0.932 (Chi-square test)

Variable [Missing]	Total (7,455)	Cancer or a malignant tumour (164)	No cancer or a malignant tumour (7,291)	*p*-value
Eat less sugar to lose/maintain weight [0]	2,140 (28.7%)	44 (26.8%)	2,096 (28.7%)	p = 0.653 (Chi-square test)
Taking more exercise to lose/maintain weight [0]	3,994 (53.6%)	66 (40.2%)	3,928 (53.9%)	p < 0.001 (Chi-square test)
Eat healthier to lose/maintain weight [0]	13 (0.174%)	0 (0%)	13 (0.178%)	p = 1 (Chi-square test)
Reduce portion to lose/maintain weight [0]	15 (0.201%)	0 (0%)	15 (0.206%)	p = 1 (Chi-square test)
Eating more protein to lose/maintain weight [0]	4 (0.054%)	0 (0%)	4 (0.055%)	p = 1 (Chi-square test)
Fasting/cutting out a meal to lose/maintain weight [0]	13 (0.174%)	0 (0%)	13 (0.178%)	p = 1 (Chi-square test)
Eat fewer carbohydrates to lose/maintain weight [0]	5 (0.067%)	0 (0%)	5 (0.069%)	p = 1 (Chi-square test)
Drink more water to lose/maintain weight [0]	6 (0.08%)	0 (0%)	6 (0.082%)	p = 1 (Chi-square test)
Eat earlier to lose/maintain weight [0]	2 (0.027%)	0 (0%)	2 (0.027%)	p = 1 (Chi-square test)
Don't eat late to lose/maintain weight [0]	4 (0.054%)	0 (0%)	4 (0.055%)	p = 1 (Chi-square test)
Slimming World to lose/maintain weight [0]	8 (0.107%)	1 (0.61%)	7 (0.096%)	p = 0.435 (Chi-square test)
Drink less alcohol/give up alcohol to lose/maintain weight [0]	3 (0.04%)	0 (0%)	3 (0.041%)	p = 1 (Chi-square test)
Use Diet App to lose/maintain weight [0]	12 (0.161%)	1 (0.61%)	11 (0.151%)	p = 0.642 (Chi-square test)
Single, never married, and never in a civil partnership [0]	2,483 (33.3%)	27 (16.5%)	2,456 (33.7%)	p < 0.001 (Chi-square test)
Married (first marriage) [0]	4,064 (54.5%)	97 (59.1%)	3,967 (54.4%)	p = 0.26 (Chi-square test)
Re-married [0]	64 (0.858%)	5 (3.05%)	59 (0.809%)	p = 0.008 (Chi-square test)
In a registered same-sex civil partnership [0]	11 (0.148%)	0 (0%)	11 (0.151%)	p = 1 (Chi-square test)
Separated [0]	241 (3.23%)	8 (4.88%)	233 (3.2%)	p = 0.326 (Chi-square test)
Divorced [0]	230 (3.09%)	6 (3.66%)	224 (3.07%)	p = 0.841 (Chi-square test)
Widowed [0]	362 (4.86%)	21 (12.8%)	341 (4.68%)	p < 0.001 (Chi-square test)
Education level [103]				p < 0.001 (Chi-square test)
- Basic - Primary education (FETAC Level 1 or 2 Cert. or equivalent). NFQ levels 1 or 2	348 (4.67%)	24 (14.7%)	324 (4.51%)	
- Basic - Lower secondary education (Junior/Inter/Group Cert, Fetac Level 3 Cert, FAS Introductory Skills, NCVA	804 (10.8%)	25 (15.3%)	779 (10.8%)	

Variable [Missing]	Total (7,455)	Cancer or a malignant tumour (164)	No cancer or a malignant tumour (7,291)	*p*-value
Foundation Cert. or equivalent. NFQ level 3				
- Intermediate - Upper secondary education (Leaving Cert. (including Applied and Vocational programs) or equivalent. NFQ levels 4 or 5. Technical or Vocational, FETAC Level 4/5 Cert., NCVA Level 1/2, FAS National Craft Cert., Teagasc Farming Cert., CERT Professional Cookery Cert. or equivalent. NFQ levels 4 or 5	1,732 (23.2%)	44 (27%)	1,688 (23.5%)	
- Advanced - Advanced Certificate / Completed Apprenticeship, FETAC Advance Cert., NCVA Level 3, FAS National Craft Cert., Teagasc Farming Cert., CERT Professional Cookery Cert. or equivalent. NFQ level 5	501 (6.72%)	7 (4.29%)	494 (6.87%)	
- Advanced - Higher Certificate, NCEA/HETAC National Cert. or equivalent. NFQ level 6	617 (8.28%)	13 (7.98%)	604 (8.4%)	
- Advanced - Ordinary Bachelor Degree or National Diploma. NFQ Level 7 and Honours Bachelor Degree/Professional qualification or both. NFQ Level 8	2,278 (30.6%)	38 (23.3%)	2,240 (31.2%)	
- Postgraduate Diploma, Masters Degree or equivalent. NFQ Level 9	971 (13%)	7 (4.29%)	964 (13.4%)	
- Doctorate (Ph.D) or higher. NFQ level 10	101 (1.35%)	5 (3.07%)	96 (1.34%)	
Limited time in the sun [0]	3,076 (41.3%)	84 (51.2%)	2,992 (41%)	p = 0.011 (Chi-square test)
Long sleeves, hat, or sunglasses for sun protection [0]	4,685 (62.8%)	118 (72%)	4,567 (62.6%)	p = 0.018 (Chi-square test)
Sunscreen with at least factor 30 [0]	4,809 (64.5%)	116 (70.7%)	4,693 (64.4%)	p = 0.109 (Chi-square test)
BMI categorised [592]				p = 0.691 (Chi-square test)
- Underweight (BMI value of less than 18.49)	116 (1.56%)	2 (1.29%)	114 (1.7%)	
- Normal (BMI value of 18.5 to 24.9)	2,726 (36.6%)	56 (36.1%)	2,670 (39.8%)	
- Overweight (BMI value of 25.0 to 29.9)	2,600 (34.9%)	60 (38.7%)	2,540 (37.9%)	
- NET Obese (BMI value of 30 or larger)	1,421 (19.1%)	37 (23.9%)	1,384 (20.6%)	
Underweight (BMI value of less than 18.49) [0]	116 (1.56%)	2 (1.22%)	114 (1.56%)	p = 0.974 (Chi-square test)

Variable [Missing]	Total (7,455)	Cancer or a malignant tumour (164)	No cancer or a malignant tumour (7,291)	*p*-value
Normal (BMI value of 18.5 to 24.9) [0]	2,726 (36.6%)	56 (34.1%)	2,670 (36.6%)	p = 0.57 (Chi-square test)
Overweight (BMI value of 25.0 to 29.9) [0]	2,600 (34.9%)	60 (36.6%)	2,540 (34.8%)	p = 0.703 (Chi-square test)
NET Obese (BMI value of 30 or larger) [0]	1,421 (19.1%)	37 (22.6%)	1,384 (19%)	p = 0.292 (Chi-square test)

Figure 2 presents the percentage of participants diagnosed with cancer and its association with BMI, tobacco use, and alcohol consumption across different Irish regions. No clear association was observed between tobacco use and cancer rates. However, the percentage of participants consuming alcohol three or more times a week was higher in regions with elevated cancer rates. Regarding BMI, the results suggest a potential association with cancer prevalence. Specifically, regions with a mean BMI in the range of 26.7 to 26.9 showed a higher percentage of cancer diagnoses. While this indicates a link within a specific BMI range, the data did not reveal a consistent trend of increasing cancer rates with increasing BMI beyond this range.

Figure 2: Discrepancies across Irish regions in cancer diagnosis by BMI, tobacco use, and alcohol consumption.

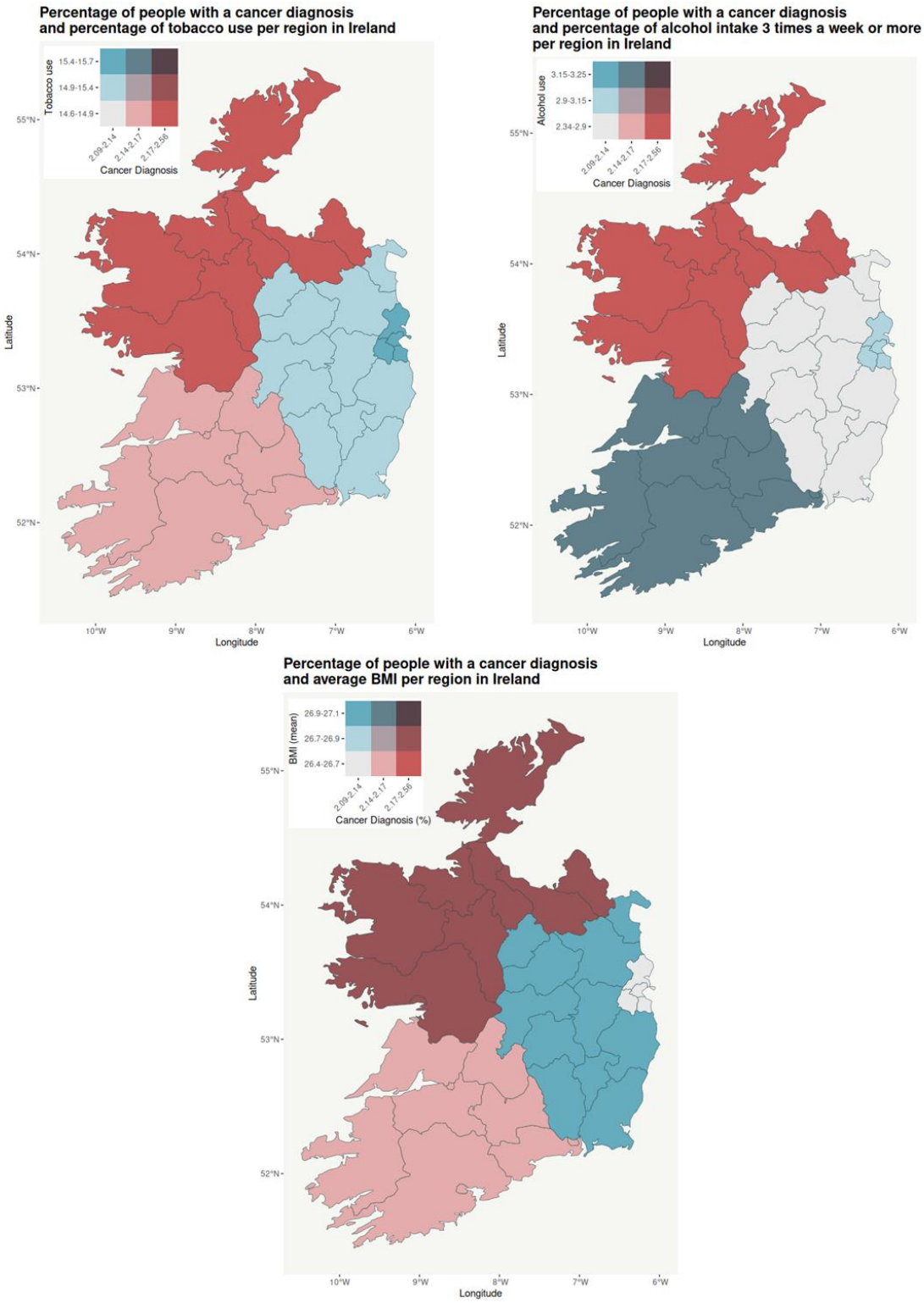


Figure 3 focuses on the discrepancies in cancer diagnosis among individuals aged between 15 and 54 across Irish regions. The map indicates that cancer prevalence in this younger age group is relatively low across the country, with subtle variations. Nonetheless, Conn/Ulster and Munster show higher cancer percentages compared to Dublin and the rest of Leinster. The results suggest that Dublin and Leinster have a younger population compared to other Irish regions and a lower cancer incidence.

Figure 3: Discrepancies in cancer diagnosis across Irish regions for individuals aged 15-54 years.

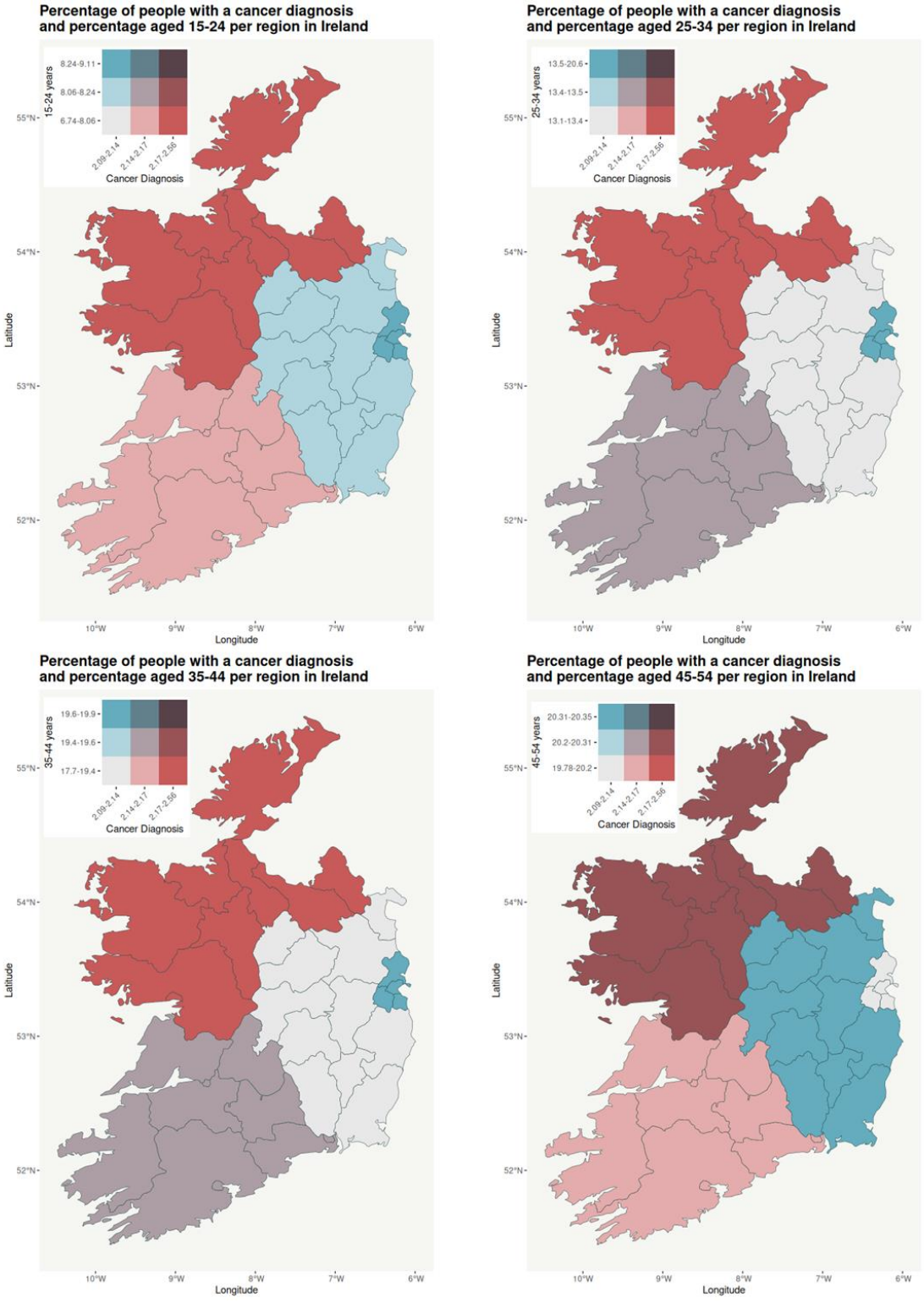


Figure 4 depicts cancer prevalence among individuals aged 55 and older across Irish regions. The colour gradients show a clear increase in cancer rates for this age group compared to younger individuals, with Conn/Ulster and Munster showing the highest percentages of diagnosed cancer cases. Dublin and Leinster present slightly lower cancer prevalence and a younger population. The visual data suggests that cancer becomes much more prevalent with age, particularly in the 75+ population of Conn/Ulster, displayed in the bottom map, which presents the highest cancer rate.

Figure 4: Discrepancies in cancer diagnosis across Irish regions for individuals aged 55 years and above.

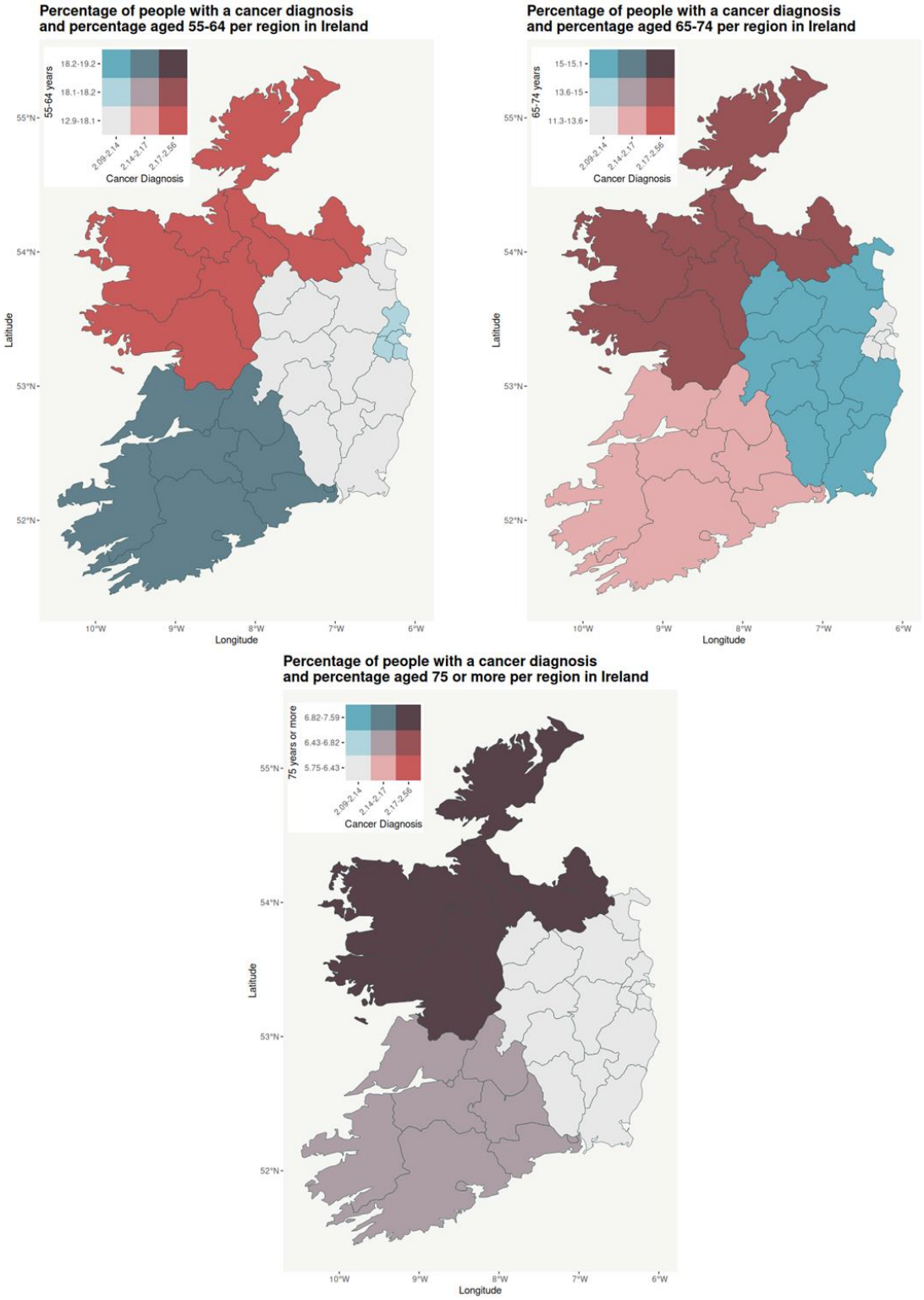


Table 3 provides predictors associated with a cancer diagnosis, focusing on age, marital status, education level, and health behaviours such as tobacco and alcohol use. Age was the most significant predictor of cancer diagnosis, with individuals aged 55-64 years having 18.3 times higher odds (95% CI: 3.96 - 326, p = 0.004) and those aged 65-74 years having 36.9 times higher odds (95% CI: 8.11 - 654, p < 0.001) compared to the 15-24 age group. Marital status was also a notable predictor, where remarried individuals presented significantly higher chances of being diagnosed with cancer (odds ratio (OR): 7.71, 95% CI: 2.54 - 19.2, p < 0.001), while married individuals presented 2.22 times higher odds (95% CI: 1.47 - 3.47, p < 0.001) compared to those never married. This significant association may partly reflect age differences, as individuals who are older and thus at greater risk for cancer are more likely to be married or remarried, whereas younger individuals tend to be single.[11] Furthermore, education level was inversely related to cancer diagnosis, with higher education levels showing a reduced risk (p < 0.001 for most). For example, compared to individuals with primary education only, those with upper secondary education had a 64.8% lower risk (OR: 0.352, 95% CI: 0.213, 0.595), and those with postgraduate degrees showed a 90.2% reduction (OR: 0.098, 95% CI: 0.039, 0.218). These results highlight the role of education in promoting healthier lifestyle choices and better healthcare access. Among lifestyle factors, limited time in the sun and long sleeves, hat or sunglasses usage for sun protection presented a lower significant correlation with cancer (OR: 1.51, 95% CI: 1.11 - 2.06, p = 0.009 and OR: 1.53, 95% CI: 1.09 - 2.18, p = 0.015, respectively). Participants willing to practice more exercise to lose or maintain weight were less likely to have a cancer diagnosis (OR: 0.577 95% CI: 0.419 - 0.788, p < 0.001). Alcohol consumption patterns and tobacco use showed mixed results. No usage of tobacco in the past was significantly associated with a reduced likelihood of cancer diagnosis (OR: 0.49, 95% CI: 0.346 - 0.698, p < 0.001), suggesting a protective effect compared to those who used tobacco daily. Regarding alcohol consumption, participants who reported only having a few sips of alcohol in their lifetime demonstrated a higher likelihood of being diagnosed with cancer compared to those who reported consuming alcohol (OR: 1.62, 95% CI: 0.986 - 2.54, p = 0.044). Moreover, those who consumed 3-4 drinks per day over the past six months presented a lower likelihood of a cancer diagnosis compared to the referent group of 1-2 drinks per day (OR: 0.572, 95% CI: 0.334 - 0.954, p = 0.036). However, it’s important to note that these findings do not imply a protective effect of higher alcohol consumption against cancer. For instance, individuals reporting lower consumption may do so because they have already been diagnosed with cancer.

Table 3: Predictive factors for cancer diagnosis, including sociodemographic and lifestyle variables.

Predictor	Cancer diagnosis
Age category: 15-24 years	1 [Referent]
Age category: 25-34 years	2.64 (0.424, 50.6) [p = 0.376]
Age category: 35-44 years	4.71 (0.913, 86.2) [p = 0.138]
Age category: 45-54 years	9.55 (2.01, 171) [p = 0.027]
Age category: 55-64 years	18.3 (3.96, 326) [p = 0.004]
Age category: 65-74 years	36.9 (8.11, 654) [p < 0.001]
Age category: 75 years and over	41.7 (8.91, 743) [p < 0.001]
Female: FALSE	1 [Referent]
Female: TRUE	0.769 (0.559, 1.05) [p = 0.102]

Predictor	Cancer diagnosis
Marital status: Single, never married, and never in a civil partnership	1 [Referent]
Marital status: Married or In a registered same-sex civil partnership	2.22 (1.47, 3.47) [p < 0.001]
Marital status: Re-married	7.71 (2.54, 19.2) [p < 0.001]
Marital status: Separated	3.12 (1.31, 6.65) [p = 0.005]
Marital status: Divorced	2.44 (0.902, 5.58) [p = 0.051]
Marital status: Widowed	5.6 (3.1, 9.99) [p < 0.001]
Region: Conn/Ulster	1 [Referent]
Region: Dublin	0.833 (0.531, 1.33) [p = 0.433]
Region: Munster	0.844 (0.527, 1.37) [p = 0.484]
Region: Rest of Leinster	0.811 (0.505, 1.32) [p = 0.388]
Number of snack foods portions: [1, 2]	1 [Referent]
Number of snack foods portions: (2, 20]	1.13 (0.63, 1.9) [p = 0.667]
Eat fewer calories to lose/maintain weight: FALSE	1 [Referent]
Eat fewer calories to lose/maintain weight: TRUE	1.02 (0.721, 1.43) [p = 0.902]
Eat less fat to lose/maintain weight: FALSE	1 [Referent]
Eat less fat to lose/maintain weight: TRUE	0.964 (0.642, 1.41) [p = 0.854]
Eat less sugar to lose/maintain weight: FALSE	1 [Referent]
Eat less sugar to lose/maintain weight: TRUE	0.909 (0.635, 1.28) [p = 0.591]
Taking more exercise to lose/maintain weight: FALSE	1 [Referent]
Taking more exercise to lose/maintain weight: TRUE	0.577 (0.419, 0.788) [p < 0.001]
Education level: Basic - Primary education (FETAC Level 1 or 2 Cert. or equivalent). NFQ levels 1 or 2	1 [Referent]
Education level: Basic - Lower secondary education (Junior/Inter/Group Cert, Fetac Level 3 Cert, FAS Introductory Skills, NCVA Foundation Cert. or equivalent. NFQ level 3	0.433 (0.243, 0.773) [p = 0.004]
Education level: Intermediate - Upper secondary education (Leaving Cert. (including Applied and Vocational programs) or equivalent. NFQ levels 4 or 5. Technical or Vocational, FETAC Level 4/5 Cert., NCVA Level 1/2, FAS National Craft Cert., Teagasc Farming Cert., CERT Professional Cookery Cert. or equivalent. NFQ levels 4 or 5	0.352 (0.213, 0.595) [p < 0.001]
Education level: Advanced - Advanced Certificate / Completed Apprenticeship, FETAC Advance Cert., NCVA Level 3, FAS National Craft Cert., Teagasc Farming Cert., CERT Professional Cookery Cert. or equivalent. NFQ level 5	0.191 (0.075, 0.427) [p < 0.001]
Education level: Advanced - Higher Certificate, NCEA/HETAC National Cert. or equivalent. NFQ level 6	0.291 (0.142, 0.569) [p < 0.001]
Education level: Advanced - Ordinary Bachelor Degree or National Diploma. NFQ Level 7 and Honours Bachelor Degree/Professional qualification or both. NFQ Level 8	0.229 (0.137, 0.391) [p < 0.001]

Predictor	Cancer diagnosis
Education level: Postgraduate diploma, Masters Degree, or equivalent. NFQ Level 9	0.098 (0.039, 0.218) [p < 0.001]
Education level: Doctorate (Ph.D) or higher. NFQ level 10	0.703 (0.232, 1.75) [p = 0.486]
Limited time in the sun: FALSE	1 [Referent]
Limited time in the sun: TRUE	1.51 (1.11, 2.06) [p = 0.009]
Long sleeves, hat, or sunglasses for sun protection: FALSE	1 [Referent]
Long sleeves, hat, or sunglasses for sun protection: TRUE	1.53 (1.09, 2.18) [p = 0.015]
Sunscreen with at least factor 30: FALSE	1 [Referent]
Sunscreen with at least factor 30: TRUE	1.34 (0.959, 1.9) [p = 0.093]
BMI categorised: Underweight (BMI value of less than 18.49)	1 [Referent]
BMI categorised: Normal (BMI value of 18.5 to 24.9)	1.2 (0.367, 7.35) [p = 0.806]
BMI categorised: Overweight (BMI value of 25.0 to 29.9)	1.35 (0.414, 8.27) [p = 0.682]
BMI categorised: NET Obese (BMI value of 30 or larger)	1.52 (0.458, 9.45) [p = 0.565]
Overweight (BMI value of 25.0 to 29.9): FALSE	1 [Referent]
Overweight (BMI value of 25.0 to 29.9): TRUE	1.08 (0.779, 1.48) [p = 0.642]
Married (first marriage): FALSE	1 [Referent]
Married (first marriage): TRUE	1.21 (0.888, 1.67) [p = 0.229]
Alcohol consumption in a lifetime: Yes	1 [Referent]
Alcohol consumption in a lifetime: Never	0.972 (0.571, 1.56) [p = 0.91]
Alcohol consumption in a lifetime: I have only had a few sips of alcohol in my lifetime	1.62 (0.986, 2.54) [p = 0.044]
Frequency of alcohol consumption (last 6 months): Daily	1 [Referent]
Frequency of alcohol consumption (last 6 months): 5-6 times a week	3.29 (0.843, 15.9) [p = 0.099]
Frequency of alcohol consumption (last 6 months): 3-4 times a week	0.867 (0.266, 3.87) [p = 0.828]
Frequency of alcohol consumption (last 6 months): Twice or once a week	0.688 (0.244, 2.88) [p = 0.538]
Frequency of alcohol consumption (last 6 months): 2-3 times a month	0.623 (0.192, 2.78) [p = 0.472]
Frequency of alcohol consumption (last 6 months): Once or less than once a month	0.802 (0.276, 3.41) [p = 0.721]
Frequency of alcohol consumption (last 6 months): I did not drink in the last 6 months, but I drank longer ago.	2.19 (0.772, 9.21) [p = 0.198]
On a typical day, how many drinks (last 6 months): 1-2	1 [Referent]
On a typical day, how many drinks (last 6 months): 3-4	0.572 (0.334, 0.954) [p = 0.036]
On a typical day, how many drinks (last 6 months): 5-6	0.728 (0.381, 1.31) [p = 0.31]
On a typical day, how many drinks (last 6 months): 7-9	0.639 (0.261, 1.35) [p = 0.278]
On a typical day, how many drinks (last 6 months): 10+	0.572 (0.247, 1.16) [p = 0.151]

Predictor	Cancer diagnosis
Frequency of consumption of 6 or more standard drinks on one occasion: More than once a week	1 [Referent]
Frequency of consumption of 6 or more standard drinks on one occasion: Once a week	1.19 (0.487, 3.05) [p = 0.709]
Frequency of consumption of 6 or more standard drinks on one occasion: 2/3 times a month	0.808 (0.243, 2.44) [p = 0.711]
Frequency of consumption of 6 or more standard drinks on one occasion: Once a month	0.626 (0.205, 1.81) [p = 0.389]
Frequency of consumption of 6 or more standard drinks on one occasion: Less than once a month	0.804 (0.33, 2.06) [p = 0.635]
Frequency of consumption of 6 or more standard drinks on one occasion: Never	1.45 (0.722, 3.34) [p = 0.333]
Who were you drinking with? (last day you had an alcoholic drink): On my own	1 [Referent]
Who were you drinking with? (last day you had an alcoholic drink): With my spouse/partner only	0.907 (0.486, 1.82) [p = 0.77]
Who were you drinking with? (last day you had an alcoholic drink): With one other person (not spouse/partner)	0.709 (0.287, 1.69) [p = 0.439]
Who were you drinking with? (last day you had an alcoholic drink): In a group of people	0.72 (0.378, 1.46) [p = 0.337]
Where were you drinking: At home (inside or outside)	1 [Referent]
Where were you drinking: In someone else’s home (inside or outside)	0.228 (0.037, 0.737) [p = 0.041]
Where were you drinking: In/outside a pub or in a nightclub	0.786 (0.464, 1.29) [p = 0.354]
Where were you drinking: In/outside a restaurant	1.18 (0.658, 2.03) [p = 0.557]
Where were you drinking: In a public place outdoors, while on the move, or somewhere else	1.37 (0.077, 6.56) [p = 0.756]
Where were you drinking: At a sport/music event, sports club/centre, or at a function/wedding	1.45 (0.235, 4.82) [p = 0.61]
Smoke tobacco products: Yes, daily	1 [Referent]
Smoke tobacco products: Yes, occasionally	0.815 (0.186, 2.55) [p = 0.751]
Smoke tobacco products: No	1.61 (0.946, 2.99) [p = 0.103]
Tobacco products (in the past): Yes, daily	1 [Referent]
Tobacco products (in the past): Yes, occasionally	0.878 (0.484, 1.51) [p = 0.652]
Tobacco products (in the past): No	0.49 (0.346, 0.698) [p < 0.001]
Tobacco habits: I have never tried e-cigarettes	1 [Referent]
Tobacco habits: I have tried e-cigarettes but do not use them (anymore)	0.842 (0.481, 1.38) [p = 0.519]
Tobacco habits: I have tried e-cigarettes and still use them daily	0.645 (0.158, 1.72) [p = 0.456]
Tobacco habits: I have tried e-cigarettes and still use them occasionally	0.563 (0.093, 1.79) [p = 0.423]

3.1.2 Irish Health Survey

The Irish Health Survey (IHS) is the Irish adaptation of the European Health Interview Survey (EHIS), conducted under Regulation (EC) No 1338/2008 of the European Parliament and the Council.[12] It enables cross-national health comparisons with other European countries, providing essential data on health status, healthcare utilisation, and health determinants. The IHS serves public health policy by offering reliable insights into the health behaviours and sociodemographic variables of the Irish population, focusing on longstanding health problems in general. The survey includes modules on health status, healthcare use, and health determinants of individuals aged 15 years or older, covering factors like gender, age, employment status, and health behaviours such as smoking, physical activity, and alcohol consumption.[13] Data collection for the 2019 IHS involved face-to-face interviews, and the final sample size of 6,759 respondents ensures a robust analysis of health trends, aiding targeted public health interventions. Unlike the paper-based 2015 IHS, the 2019 survey used Computer-Assisted Personal Interviewing (CAPI) methods, enhancing the accuracy and consistency of the responses.

Table 4 presents the distribution of health and sociodemographic variables for a total sample of 6,759 participants. Among the age groups, 42.6% of participants were aged 40 to 64, 29% were aged 15 to 39, and 28.4% were aged 65 or older. The gender distribution showed a slight skew toward females, who comprised 54% of the sample. Regarding BMI, 38.4% of participants fell into the normal weight category (BMI: 18.5–24.9), while 34.5% were classified as overweight (BMI: 24.9–29.9), and 24.2% were categorised as obese (BMI: ≥29.9). A notable portion of the sample (43.9%) was classified as disadvantaged or very disadvantaged according to the deprivation index, while only 17.5% fell into the very affluent category. The deprivation index is a composite measure used to capture relative socioeconomic disadvantage in a population, considering factors such as income, employment, education, and housing quality.

In terms of employment status, over half of the participants (50.8%) were working for payment or profit, with a significant number retired (24.2%) or unable to work due to disability (5.07%). Regarding general health, 42.7% of participants reported “very good” health, while 30.7% reported having longstanding illnesses. Additionally, 22.37% registered some level of activity limitation due to health problems.

Key lifestyle factors revealed important health behaviours among the population. Alcohol consumption varied, with 28.3% drinking 1 to 2 days per week, while 17.2% reported never having consumed alcohol or only a few sips during their lifetime. Concerning nutrition, 65.9% consumed fruit daily, and 71.7% ate vegetables, though 57% indicated that they never consumed sugary soft drinks. Regarding mental health, depression was diagnosed in 4.59% of participants. Furthermore, physical activity levels were low, with 54.5% never engaging in sports or fitness activities and only 7.06% exercising daily. Lastly, 14.4% of participants were daily smokers, whereas 81.8% did not smoke, reflecting a robust anti-smoking trend. Overall, the population reflects a diverse sample in terms of health outcomes, socioeconomic status, and health behaviours.

Table 4: Health, sociodemographics, and lifestyle factors among participants of the Irish Health Survey.

Variable [Missing]	Total (6,759)
Age category [0]	
- 15 - 39	1,961 (29%)
- 40 - 64	2,880 (42.6%)
- 65+	1,918 (28.4%)
Gender [0]	
- Female	3,651 (54%)
- Male	3,108 (46%)
Body Mass Index (BMI) [8]	
- <18.5	187 (2.77%)
- 18.5 - 24.9	2,593 (38.4%)
- 24.9 - 29.9	2,330 (34.5%)
- 29.9 - 34.9	929 (13.7%)
- 34.9+	712 (10.5%)
Weight range (kgs) [8]	
- 40-59	1,119 (16.6%)
- 59.1 - 79	3,293 (48.7%)
- 79.1 - 99	1,853 (27.4%)
- 99.1+	486 (7.19%)
Height range (cms) [1]	
- <=149	746 (11%)
- >= 179.1	1,179 (17.4%)
- 149.1 - 179	4,833 (71.5%)
Deprivation index [8]	
- Very Disadvantaged	1,567 (23.2%)
- Disadvantaged	1,399 (20.7%)
- Average	1,266 (18.7%)
- Affluent	1,339 (19.8%)
- Very Affluent	1,180 (17.5%)
Degree of urbanisation [0]	
- Cities	2,231 (33%)

Variable [Missing]	Total (6,759)
- Rural	3,270 (48.4%)
- Towns and Suburbs	1,258 (18.6%)
Nationality [1]	
- Irish National	6,114 (90.5%)
- Non-National	644 (9.53%)
Employment status [6]	
- Working for payment or profit	3,435 (50.8%)
- Looking for first regular job	12 (0.178%)
- Unemployed	209 (3.09%)
- A student or pupil	392 (5.8%)
- Retired from employment	1,633 (24.2%)
- Unable to work due to permanent sickness or disability	343 (5.07%)
- Looking after a family home	684 (10.1%)
- Other	45 (0.666%)
Absent from work for reasons of health problems (in the past 12 months) [2,746]	672 (9.94%)
Number of days absent from work for reasons of health problems (in the past 12 months) [6,098]	
- <5	191 (2.83%)
- 10+	319 (4.72%)
- 5-9	151 (2.23%)
None (In the past 12 months, have you been absent from work for reasons of health problems?) [394]	6,365 (94.2%)
How is the health in general [6]	
- Very good	2,888 (42.7%)
- Good	2,541 (37.6%)
- Fair	1,011 (15%)
- Bad	234 (3.46%)
- Very bad	79 (1.17%)
Longstanding illness or longstanding health problem [9]	2,074 (30.7%)
Limited because of a health problem in activities people usually do [7]	
- Severely limited	424 (6.27%)
- Limited but not severely	1,088 (16.1%)
- Not limited at all	5,240 (77.5%)
Limited for at least the past 6 months [5,247]	1,348 (19.9%)

Variable [Missing]	Total (6,759)
Limited - Preparing meals [3,864]	
- No difficulty	2,772 (41%)
- Some difficulty	123 (1.82%)
Limited - Using the telephone [3,842]	
- No difficulty	2,917 (43.2%)
Limited - Shopping [3,953]	
- No difficulty	2,634 (39%)
- Some difficulty	172 (2.54%)
Limited - Managing medication [3,860]	
- No difficulty	2,811 (41.6%)
- Some difficulty	88 (1.3%)
Limited - Light housework [3,893]	
- No difficulty	2,699 (39.9%)
- Some difficulty	167 (2.47%)
Limited - Occasional heavy housework [4,010]	
- No difficulty	2,093 (31%)
- Some difficulty	435 (6.44%)
- A lot of difficulty	221 (3.27%)
Limited - Taking care of finances and everyday administrative tasks [3,851]	
- No difficulty	2,780 (41.1%)
- Some difficulty	128 (1.89%)
Usually, have help with any of these activities [5,789]	
- No	434 (6.42%)
- Yes, with at least one activity	536 (7.93%)
Need more help [6,226]	
- No	323 (4.78%)
- Yes, with at least one activity	210 (3.11%)
Need help [6,325]	
- No	341 (5.05%)
- Yes, with at least one activity	93 (1.38%)
How much bodily pain (during the past 4 weeks) [6]	
- None	3,608 (53.4%)

Variable [Missing]	Total (6,759)
- Very mild	785 (11.6%)
- Mild	896 (13.3%)
- Moderate	854 (12.6%)
- Severe	479 (7.09%)
- Very Severe	131 (1.94%)
How much did pain interfere with your normal life (during the past 4 weeks) [6]	
- Not at all	4,603 (68.1%)
- A little bit	999 (14.8%)
- Moderately	490 (7.25%)
- Quite a bit	450 (6.66%)
- Extremely	211 (3.12%)
How often had an alcoholic drink of any kind (in the past 12 months) [34]	
- Every day or almost	213 (3.15%)
- 5-6 days a week	94 (1.39%)
- 3-4 days a week	506 (7.49%)
- 1-2 days a week	1,914 (28.3%)
- 2-3 days a month	729 (10.8%)
- Once a month	642 (9.5%)
- Less than once a month	848 (12.5%)
- Not in the past 12 months, as I no longer drink alcohol	614 (9.08%)
- Never, or only a few sips or tries, in my whole life	1,165 (17.2%)
Thinking of Monday to Thursday, how many of these 4 days do you usually drink alcohol [4,331]	
- On 2 of the 4 days	188 (2.78%)
- On 1 of the 4 days	573 (8.48%)
- None of the days	1,667 (24.7%)
From Monday to Thursday, how many drinks do you have on average on such a day when you drink alcohol? [6,087]	
- 4-5 drinks a day	160 (2.37%)
- 3 drinks a day	192 (2.84%)
- 2 drinks a day	320 (4.73%)
Thinking of Friday to Sunday, how many of these 3 days do you usually drink alcohol? [1,622]	
- On all 3 days	484 (7.16%)

Variable [Missing]	Total (6,759)
- On 2 of the 3 days	1,220 (18.1%)
- On 1 of the 3 days	2,253 (33.3%)
- On none of the days	1,180 (17.5%)
From Friday to Sunday, how many drinks do you have on average on such a day when you drink alcohol? [4,485]	
- 6-9 drinks a day	311 (4.6%)
- 4-5 drinks a day	733 (10.8%)
- 3 drinks a day	653 (9.66%)
- 2 drinks a day	577 (8.54%)
How often have you had six or more drinks containing alcohol on one occasion? (in the past 12 months) [1,925]	
- 1-2 days a week	309 (4.57%)
- 2-3 days a month	330 (4.88%)
- Once a month	749 (11.1%)
- Less than once a month	1,490 (22%)
- Not in the past 12 months	1,277 (18.9%)
- Never in my whole life	679 (10%)
How often eat fruit [9]	
- Once or more a day	4,451 (65.9%)
- 4 to 6 times a week	943 (14%)
- 1 to 3 times a week	792 (11.7%)
- Less than once a week	361 (5.34%)
- Never	203 (3%)
How many portions of fruit [2,315]	
- 1	974 (14.4%)
- 2	1,594 (23.6%)
- 3	1,129 (16.7%)
- 4	359 (5.31%)
- 5	292 (4.32%)
- 6	52 (0.769%)
- 7	14 (0.207%)
- 8+	30 (0.444%)
How often eat vegetables or salad [8]	
- Once or more a day	4,849 (71.7%)

Variable [Missing]	Total (6,759)
- 4 to 6 times a week	1,042 (15.4%)
- 1 to 3 times a week	613 (9.07%)
- Less than once a week	148 (2.19%)
- Never	99 (1.46%)
How many portions of vegetables or salad (Mean/SD) [1,917]	2.55 (+-2.05)
How often do you drink 100% pure fruit or vegetable juice [12]	
- Once or more a day	888 (13.1%)
- 4 to 6 times a week	394 (5.83%)
- 1 to 3 times a week	707 (10.5%)
- Less than once a week	819 (12.1%)
- Never	3,939 (58.3%)
How often do you drink sugared soft drinks [9]	
- Once or more a day	530 (7.84%)
- 4 to 6 times a week	291 (4.31%)
- 1 to 3 times a week	886 (13.1%)
- Less than once a week	1,188 (17.6%)
- Never	3,855 (57%)
Depression diagnosis [6,449]	310 (4.59%)
Working - physical activity level [21]	
- Mostly sitting or standing	2,229 (33%)
- Mostly walking or tasks of moderate physical effort	1,826 (27%)
- Mostly heavy labour or physically demanding work	458 (6.78%)
- Quite a bit	2,225 (32.9%)
How many days do you walk for at least 10 minutes continuously in order to get to and from places [37]	
- 1 day	182 (2.69%)
- 2 days	295 (4.36%)
- 3 days	468 (6.92%)
- 4 days	476 (7.04%)
- 5 days	819 (12.1%)
- 6 days	268 (3.97%)
- 7 days	2,910 (43.1%)
- I never carry out such a physical activity	1,304 (19.3%)

Variable [Missing]	Total (6,759)
How much time spend walking [1,358]	
- 10 - 29 minutes per day	2,736 (40.5%)
- 30 - 59 minutes per day	1,751 (25.9%)
- 1 hour to less than 2 hours per day	629 (9.31%)
- 2 hours to less than 3 hours per day	147 (2.17%)
- 3 hours or more per day	138 (2.04%)
How many days do you BICYCLE for at least 10 minutes continuously to get to and from places [35]	
- 1-3 days	499 (7.38%)
- 4-7 days	267 (3.95%)
- I never carry out such a physical activity	5,958 (88.1%)
How much time do you spend bicycling [6,180]	
- 10 - 29 minutes per day	330 (4.88%)
- 30 - 59 minutes per day	249 (3.68%)
How many days do you carry out sports, fitness, or recreational physical activities [22]	
- 1 day	400 (5.92%)
- 2 days	520 (7.69%)
- 3 days	697 (10.3%)
- 4 days	430 (6.36%)
- 5 days	414 (6.13%)
- 6 days	115 (1.7%)
- 7 days	477 (7.06%)
- I never carry out such a physical activity	3,684 (54.5%)
How much time in total do you spend on sports, fitness, or recreational physical activities [4,111]	
- 0.0 - 1.0 hours	313 (4.63%)
- 1.1 - 2.0 hours	429 (6.35%)
- 2.1 - 3.0 hours	487 (7.21%)
- 3.1 - 4.0 hours	324 (4.79%)
- 4.1 - 5.0 hours	279 (4.13%)
- 5.1 - 6.0 hours	233 (3.45%)
- 6.1+	583 (8.63%)
How many days do you carry out physical activities specifically designed to strengthen your muscles [21]	

Variable [Missing]	Total (6,759)
- 1-3 days	1,053 (15.6%)
- 4-6 days	353 (5.22%)
- 7 days	149 (2.2%)
- I never carry out such a physical activity	5,183 (76.7%)
How much time is spent sitting and reclining on a typical day [47]	
- Less than 4 hours	3,362 (49.7%)
- 4 hours to less than 6 hours.	1,800 (26.6%)
- 6 hours to less than 8 hours	840 (12.4%)
- 8 hours to less than 10 hours	437 (6.47%)
- 10 hours to less than 12 hours	167 (2.47%)
- 12 hours or more	106 (1.57%)
Smoke any tobacco products [12]	
- Yes, daily	971 (14.4%)
- Yes, occasionally	245 (3.62%)
- Not at all	5,531 (81.8%)
How many cigarettes do you smoke each day [5,807]	
- <10	232 (3.43%)
- 10-19	401 (5.93%)
- 20+	319 (4.72%)
Smoke tobacco daily, or almost daily, for at least one year [984]	1,947 (28.8%)
For how many years did you smoke daily [3,869]	
- <5	333 (4.93%)
- 10-20	725 (10.7%)
- 20+	1,484 (22%)
- 5-10	348 (5.15%)
Exposure to tobacco smoke indoors [12]	
- Every day	639 (9.45%)
- At least once a week (but not every day)	170 (2.52%)
- Less than once a week	227 (3.36%)
- Never or almost never	5,711 (84.5%)
Currently use electronic cigarettes or similar electronic devices [311]	
- No, but former vaping	114 (1.69%)

Variable [Missing]	Total (6,759)
- Never vaping	6,334 (93.7%)
Smoke hand-rolled cigarettes each day [6,480]	279 (4.13%)
Smoke manufactured cigarettes each day [6,053]	706 (10.4%)

Table 5 presents the predictors associated with longstanding illness or health problems using logistic regression models. Age emerged as a significant predictor, with participants aged 65+ having 4.48 times higher odds of reporting a chronic health condition in comparison to those aged 15–39 (95% CI: 3.86–5.21, $p < 0.001$). Similarly, participants aged 40–64 showed more than double the odds compared to the younger group (OR: 2.23, 95% CI: 1.93–2.57, $p < 0.001$). BMI also influenced the likelihood of developing chronic conditions, with individuals with a BMI ≥34.9 showing elevated odds (OR: 1.57, 95% CI: 1.11–2.23, $p = 0.011$) compared to those with a BMI <18.5. Participants weighing 99.1 kg or more (OR: 1.49, 95% CI: 1.2–1.87, $p < 0.001$) had a significantly greater likelihood of reporting chronic health conditions compared to those weighing 40-59 kg. In contrast, no significant associations were observed for participants in the other categories relative to the referent group. Additionally, individuals taller than or equal to 171.1 cm (OR: 0.652, 95% CI: 0.534-0.795, $p < 0.001$) were significantly less likely to report chronic health issues, in contrast to those under 149 cm. No significant associations were observed for the other height ranges when compared to the referent group.

Socioeconomic status was also a significant factor, with those from the “Very Affluent” group showing 39% lower odds of developing persistent medical issues (OR: 0.609, 95% CI: 0.515-0.719, $p < 0.001$) compared to the “Very Disadvantaged” group. Employment status influenced health as well, particularly among retirees (OR: 4.02, 95% CI: 3.53-4.58, $p < 0.001$) and individuals unable to work due to disability or illness (OR: 26.9, 95% CI: 19.8-37.4, $p < 0.001$) compared to participants in paid employment.

Alcohol consumption over the past year was associated with a reduced likelihood of chronic conditions when consumed 1-2 days a week (OR: 0.534, 95% CI: 0.398-0.719, $p < 0.001$), 2-3 days a month (OR: 0.504, 95% CI: 0.365-0.699, $p < 0.001$), or once a month (OR: 0.482, 95% CI: 0.346-0.673, $p < 0.001$), compared to daily or near-daily drinking. Dietary habits also played a role; those who never ate fruit had 2.05 times higher odds of developing enduring health concerns (95% CI: 1.53-2.71, $p < 0.001$) in comparison to those who ate fruit daily. Additionally, consuming sugar-sweetened soft drinks 4-6 times a week (OR: 0.72, 95% CI: 0.517-0.994, $p = 0.048$) or 1-3 times a week (OR: 0.628, 95% CI: 0.492-0.803, $p < 0.001$) was linked to lower odds of developing chronic medical issues when compared to daily consumption.

Physical activity showed protective effects. Walking four or more days a week for at least 10 minutes reduced the odds of developing a chronic condition, with declines seen for those walking 4 days (OR: 0.569, 95% CI: 0.395-0.824, $p = 0.003$), 5 days (OR: 0.591, 95% CI: 0.421-0.833, $p = 0.002$), 6 days (OR: 0.633, 95% CI: 0.422-0.951, $p = 0.028$), or 7 days (OR: 0.627, 95% CI: 0.46-0.862, $p = 0.004$) contrasting to participants walking once a week. Furthermore, those who did not engage in walking for at least 10 minutes continuously had 1.51 times higher odds of developing a

persistent health problem (95% CI: 1.1-2.09, p = 0.012). Walking duration also mattered, with diminished odds of chronic illness for individuals walking 30-59 minutes per day (OR: 0.676, 95% CI: 0.589-0.775, p < 0.001), 1-2 hours (OR: 0.609, 95% CI: 0.493-0.747, p < 0.001), and 3+ hours (OR: 0.541, 95% CI: 0.346-0.817, p = 0.005). Conversely, prolonged sitting (4+ hours) on a typical day increased the likelihood of enduring health conditions, with participants sitting 12+ hours showing a significantly elevated risk (OR: 10.6, 95% CI: 1.82-2.67, p < 0.001) compared to those sitting less than 4 hours.

Finally, smoking tobacco daily or almost daily for at least a year had a significant impact on the development of chronic medical issues (OR: 1.57, 95% CI: 1.4-1.77, p < 0.001) compared to non-daily smokers. These findings suggest that targeting lifestyle factors such as diet, physical activity, and pain management, particularly among disadvantaged populations, could help alleviate the burden of chronic health conditions.

Table 5: Predictive factors for longstanding illness or health problems.

Predictor	Longstanding illness or longstanding health problem
Age category: 15 - 39	1 [Referent]
Age category: 40 - 64	2.23 (1.93, 2.57) [p < 0.001]
Age category: 65+	4.48 (3.86, 5.21) [p < 0.001]
Gender: Female	1 [Referent]
Gender: Male	0.953 (0.859, 1.06) [p = 0.367]
Body Mass Index (BMI): <18.5	1 [Referent]
Body Mass Index (BMI): 18.5 - 24.9	0.778 (0.566, 1.08) [p = 0.128]
Body Mass Index (BMI): 24.9 - 29.9	1.06 (0.769, 1.47) [p = 0.74]
Body Mass Index (BMI): 29.9 - 34.9	1.26 (0.902, 1.78) [p = 0.181]
Body Mass Index (BMI): 34.9+	1.57 (1.11, 2.23) [p = 0.011]
Weight range (kgs): 40-59	1 [Referent]
Weight range (kgs): 59.1 - 79	0.944 (0.814, 1.1) [p = 0.444]
Weight range (kgs): 79.1 - 99	1.1 (0.936, 1.29) [p = 0.252]
Weight range (kgs): 99.1+	1.49 (1.2, 1.87) [p < 0.001]
Height range (cms): <=149	1 [Referent]
Height range (cms): >= 179.1	0.652 (0.534, 0.795) [p < 0.001]
Height range (cms): 149.1 - 179	0.851 (0.724, 1) [p = 0.051]
Deprivation index: Very Disadvantaged	1 [Referent]
Deprivation index: Disadvantaged	0.9 (0.773, 1.05) [p = 0.171]
Deprivation index: Average	0.713 (0.608, 0.836) [p < 0.001]
Deprivation index: Affluent	0.663 (0.566, 0.777) [p < 0.001]
Deprivation index: Very Affluent	0.609 (0.515, 0.719) [p < 0.001]
Degree of urbanisation: Cities	1 [Referent]
Degree of urbanisation: Rural	0.932 (0.83, 1.05) [p = 0.237]
Degree of urbanisation: Towns and Suburbs	0.863 (0.742, 1) [p = 0.057]
Employment status: Working for payment or profit	1 [Referent]
Employment status: Looking for first regular job	0.879 (0.135, 3.34) [p = 0.868]
Employment status: Unemployed	1.53 (1.1, 2.1) [p = 0.009]
Employment status: A student or pupil	0.938 (0.709, 1.23) [p = 0.649]
Employment status: Retired from employment	4.02 (3.53, 4.58) [p < 0.001]
Employment status: Unable to work due to permanent sickness or disability	26.9 (19.8, 37.4) [p < 0.001]

Predictor	Longstanding illness or longstanding health problem
Employment status: Looking after family home	2.07 (1.73, 2.49) [p < 0.001]
Employment status: Other	2.93 (1.58, 5.31) [p < 0.001]
Absent from work for reasons of health problems (in the past 12 months): FALSE	1 [Referent]
Absent from work for reasons of health problems (in the past 12 months): TRUE	2.46 (2.05, 2.95) [p < 0.001]
How is your health in general: Very good	1 [Referent]
How is your health in general: Good	4.17 (3.59, 4.86) [p < 0.001]
How is the health in general: Fair	27.4 (22.7, 33.1) [p < 0.001]
How is the health in general: Bad	142 (85.6, 253) [p < 0.001]
How is the health in general: Very bad	182 (74.8, 600) [p < 0.001]
Limited because of a health problem in activities people usually do: Severely limited	1 [Referent]
Limited because of a health problem in activities people usually do: Limited but not severely	0.404 (0.278, 0.573) [p < 0.001]
Limited because of a health problem in activities people usually do: Not limited at all	0.019 (0.013, 0.026) [p < 0.001]
How much bodily pain (during the past 4 weeks): None	1 [Referent]
How much bodily pain (during the past 4 weeks): Very mild	2.09 (1.75, 2.48) [p < 0.001]
How much bodily pain (during the past 4 weeks): Mild	2.83 (2.4, 3.32) [p < 0.001]
How much bodily pain (during the past 4 weeks): Moderate	5.92 (5.05, 6.96) [p < 0.001]
How much bodily pain (during the past 4 weeks): Severe	11.9 (9.65, 14.8) [p < 0.001]
How much bodily pain (during the past 4 weeks): Very Severe	17.2 (11.4, 26.6) [p < 0.001]
How much did pain interfere with your normal life (during the past 4 weeks): Not at all	1 [Referent]
How much did pain interfere with your normal life (during the past 4 weeks): A little bit	2.69 (2.33, 3.12) [p < 0.001]
How much did pain interfere with your normal life (during the past 4 weeks): Moderately	6.22 (5.12, 7.57) [p < 0.001]
How much did pain interfere with your normal life (during the past 4 weeks): Quite a bit	10.5 (8.43, 13) [p < 0.001]
How much did pain interfere with your normal life (during the past 4 weeks): Extremely	13.5 (9.81, 18.9) [p < 0.001]
How often have you had an alcoholic drink of any kind (in the past 12 months): Every day or almost	1 [Referent]
How often have you had an alcoholic drink of any kind (in the past 12 months): 5-6 days a week	0.815 (0.484, 1.35) [p = 0.433]
How often have you had an alcoholic drink of any kind (in the past 12 months): 3-4 days a week	0.726 (0.52, 1.02) [p = 0.061]
How often have you had an alcoholic drink of any kind (in the past 12 months): 1-2 days a week	0.534 (0.398, 0.719) [p < 0.001]
How often have you had an alcoholic drink of any kind (in the past 12 months): 2-3 days a month	0.504 (0.365, 0.699) [p < 0.001]

Predictor	Longstanding illness or longstanding health problem
How often have you had an alcoholic drink of any kind (in the past 12 months): Once a month	0.482 (0.346, 0.673) [p < 0.001]
How often had an alcoholic drink of any kind (in the past 12 months): Less than once a month	0.77 (0.564, 1.05) [p = 0.1]
How often have you had an alcoholic drink of any kind (in the past 12 months): Not in the past 12 months, as I no longer drink alcohol	1.54 (1.12, 2.12) [p = 0.008]
How often have you had an alcoholic drink of any kind (in the past 12 months): Never, or only a few sips or tries, in my whole life	1 (0.745, 1.36) [p = 0.974]
Thinking of Friday to Sunday, how many of these 3 days do you usually drink alcohol?: On all 3 days	1 [Referent]
Thinking of Friday to Sunday, how many of these 3 days do you usually drink alcohol?: On 2 of the 3 days	0.652 (0.519, 0.819) [p < 0.001]
Thinking of Friday to Sunday, how many of these 3 days do you usually drink alcohol?: On 1 of the 3 days	0.613 (0.497, 0.758) [p < 0.001]
Thinking of Friday to Sunday, how many of these 3 days do you usually drink alcohol?: On none of the days	0.968 (0.775, 1.21) [p = 0.775]
How often have you had six or more drinks containing alcohol on one occasion? (in the past 12 months): 1-2 days a week	1 [Referent]
How often have you had six or more drinks containing alcohol on one occasion? (in the past 12 months): 2-3 days a month	0.697 (0.483, 1) [p = 0.052]
How often have you had six or more drinks containing alcohol on one occasion? (in the past 12 months): Once a month	0.688 (0.507, 0.936) [p = 0.016]
How often have you had six or more drinks containing alcohol on one occasion? (in the past 12 months): Less than once a month	0.843 (0.642, 1.12) [p = 0.227]
How often have you had six or more drinks containing alcohol on one occasion? (in the past 12 months): Not in the past 12 months	1.32 (1, 1.74) [p = 0.049]
How often have you had six or more drinks containing alcohol on one occasion? (in the past 12 months): Never in my whole life	1.09 (0.813, 1.48) [p = 0.553]
How often do you eat fruit: Once or more a day	1 [Referent]
How often do you eat fruit: 4 to 6 times a week	1.03 (0.883, 1.2) [p = 0.694]
How often do you eat fruit: 1 to 3 times a week	1.29 (1.1, 1.51) [p = 0.002]
How often eat fruit: Less than once a week	1.63 (1.31, 2.03) [p < 0.001]
How often do you eat fruit: Never	2.04 (1.53, 2.71) [p < 0.001]
How many portions of fruit: 1	1 [Referent]
How many portions of fruit: 2	0.986 (0.827, 1.18) [p = 0.871]
How many portions of fruit: 3	0.901 (0.745, 1.09) [p = 0.285]
How many portions of fruit: 4	1.09 (0.839, 1.42) [p = 0.507]
How many portions of fruit: 5	0.99 (0.74, 1.32) [p = 0.944]
How many portions of fruit: 6	1.39 (0.763, 2.46) [p = 0.269]
How many portions of fruit: 7	0.964 (0.263, 2.91) [p = 0.951]
How many portions of fruit: 8+	0.877 (0.363, 1.92) [p = 0.753]
How often do you eat vegetables or salad: Once or more a day	1 [Referent]

Predictor	Longstanding illness or longstanding health problem
How often do you eat vegetables or salad: 4 to 6 times a week	1.02 (0.883, 1.18) [p = 0.772]
How often eat vegetables or salad: 1 to 3 times a week	1.15 (0.956, 1.37) [p = 0.138]
How often do you eat vegetables or salad: Less than once a week	1.33 (0.94, 1.86) [p = 0.102]
How often do you eat vegetables or salad: Never	1.21 (0.787, 1.83) [p = 0.374]
How many portions of vegetables or salad: [1, 2]	1 [Referent]
How many portions of vegetables or salad: (2, 56]	0.8 (0.705, 0.908) [p < 0.001]
How often do you drink 100% pure fruit or vegetable juice: Once or more a day	1 [Referent]
How often do you drink 100% pure fruit or vegetable juice: 4 to 6 times a week	0.757 (0.575, 0.991) [p = 0.045]
How often do you drink 100% pure fruit or vegetable juice: 1 to 3 times a week	0.852 (0.682, 1.06) [p = 0.155]
How often do you drink 100% pure fruit or vegetable juice: Less than once a week	0.893 (0.723, 1.1) [p = 0.291]
How often drink 100% pure fruit or vegetable juice: Never	1.18 (1.01, 1.38) [p = 0.041]
How often do you drink sugared soft drinks: Once or more a day	1 [Referent]
How often do you drink sugared soft drinks: 4 to 6 times a week	0.72 (0.517, 0.994) [p = 0.048]
How often do you drink sugared soft drinks: 1 to 3 times a week	0.628 (0.492, 0.803) [p < 0.001]
How often do you drink sugared soft drinks: Less than once a week	0.828 (0.661, 1.04) [p = 0.1]
How often do you drink sugared soft drinks: Never	1.2 (0.988, 1.46) [p = 0.068]
Working - physical activity level: Mostly sitting or standing	1 [Referent]
Working - physical activity level: Mostly walking or tasks of moderate physical effort	1.01 (0.871, 1.16) [p = 0.924]
Working - physical activity level: Mostly heavy labor or physically demanding work	0.616 (0.468, 0.8) [p < 0.001]
Working - physical activity level: Quite a bit	2.73 (2.4, 3.1) [p < 0.001]
How many days do you walk for at least 10 minutes continuously in order to get to and from places: 1 day	1 [Referent]
How many days do you walk for at least 10 minutes continuously in order to get to and from places: 2 days	0.943 (0.642, 1.39) [p = 0.765]
How many days do you walk for at least 10 minutes continuously in order to get to and from places: 3 days	0.737 (0.514, 1.06) [p = 0.099]
How many days do you walk for at least 10 minutes continuously in order to get to and from places: 4 days	0.569 (0.395, 0.824) [p = 0.003]
How many days do you walk for at least 10 minutes continuously in order to get to and from places: 5 days	0.591 (0.421, 0.833) [p = 0.002]
How many days do you walk for at least 10 minutes continuously in order to get to and from places: 6 days	0.633 (0.422, 0.951) [p = 0.028]
How many days do you walk for at least 10 minutes continuously in order to get to and from places: 7 days	0.627 (0.46, 0.862) [p = 0.004]
How many days do you walk for at least 10 minutes continuously in order to get to and from places? I have never carried out such a physical activity	1.51 (1.1, 2.09) [p = 0.012]

Predictor	Longstanding illness or longstanding health problem
How much time spent walking: 10 - 29 minutes per day	1 [Referent]
How much time spent walking: 30 - 59 minutes per day	0.676 (0.589, 0.775) [p < 0.001]
How much time spend walking: 1 hour to less than 2 hours per day	0.609 (0.493, 0.747) [p < 0.001]
How much time spent walking: 2 hours to less than 3 hours per day	0.919 (0.632, 1.31) [p = 0.65]
How much time spent walking : 3 hours or more per day	0.541 (0.346, 0.817) [p = 0.005]
How many days do you carry out sports, fitness, or recreational physical activities: 1 day	1 [Referent]
How many days do you carry out sports, fitness, or recreational physical activities: 2 days	0.746 (0.543, 1.02) [p = 0.069]
How many days do you carry out sports, fitness, or recreational physical activities: 3 days	0.845 (0.632, 1.13) [p = 0.259]
How many days do you carry out sports, fitness, or recreational physical activities: 4 days	0.77 (0.553, 1.07) [p = 0.119]
How many days do you carry out sports, fitness, or recreational physical activities: 5 days	0.883 (0.637, 1.22) [p = 0.454]
How many days do you carry out sports, fitness, or recreational physical activities: 6 days	1.1 (0.678, 1.76) [p = 0.687]
How many days do you carry out sports, fitness, or recreational physical activities: 7 days	1.28 (0.951, 1.74) [p = 0.104]
How many days do you carry out sports, fitness, or recreational physical activities: I never carry out such a physical activity	1.87 (1.48, 2.38) [p < 0.001]
How much time spent sitting and reclining on a typical day: Less than 4 hours	1 [Referent]
How much time spent sitting and reclining on a typical day: 4 hours to less than 6 hours.	1.55 (1.37, 1.76) [p < 0.001]
How much time spent sitting and reclining on a typical day: 6 hours to less than 8 hours	1.87 (1.59, 2.19) [p < 0.001]
How much time spent sitting and reclining on a typical day: 8 hours to less than 10 hours	1.74 (1.41, 2.15) [p < 0.001]
How much time spent sitting and reclining on a typical day: 10 hours to less than 12 hours	3.1 (2.26, 4.24) [p < 0.001]
How much time spent sitting and reclining on a typical day: 12 hours or more	10.6 (6.77, 17.1) [p < 0.001]
physicalactivity_bicycle_recat: 1-7 days	1 [Referent]
physicalactivity_bicycle_recat: I never carry out such a physical activity	2.2 (1.82, 2.67) [p < 0.001]
How many days do you carry out physical activities specifically designed to strengthen your muscles: 1-3 days	1 [Referent]
How many days do you carry out physical activities specifically designed to strengthen your muscles: 4-6 days	1.11 (0.823, 1.49) [p = 0.487]
How many days do you carry out physical activities specifically designed to strengthen your muscles: 7 days	1.54 (1.03, 2.25) [p = 0.031]
How many days do you carry out physical activities specifically designed to strengthen your muscles: I never carry out such a physical activity	2.05 (1.75, 2.42) [p < 0.001]
Smoke tobacco daily, or almost daily, for at least one year: FALSE	1 [Referent]

Predictor	Longstanding illness or longstanding health problem
Smoke tobacco daily, or almost daily, for at least one year: TRUE	1.57 (1.4, 1.77) [p < 0.001]
Exposure to tobacco smoke indoors: Every day	1 [Referent]
Exposure to tobacco smoke indoors: At least once a week (but not every day)	0.627 (0.433, 0.898) [p = 0.012]
Exposure to tobacco smoke indoors: Less than once a week	0.587 (0.421, 0.812) [p = 0.001]
Exposure to tobacco smoke indoors: Never or almost never	0.62 (0.524, 0.733) [p < 0.001]

3.1.3 EU statistics on income and living conditions (EU-SILC)

The European Union Statistics on Income and Living Conditions (EU-SILC) survey, managed by Eurostat, is a pivotal resource for collecting harmonised data across EU member countries. It provides valuable, comparable insights into social, economic, and health indicators, enabling a clear view of income distribution, poverty, and overall living conditions across the EU. Recent updates have standardised core variables to enhance cross-country comparability, solidifying EU-SILC’s essential role in monitoring social cohesion, poverty, and exclusion for EU and national policymakers.[14]

The 2022 Austrian EU-SILC dataset comprises 10,191 participants, with data segmented into those with chronic illnesses (3,810 participants) and those without (6,381 participants). This dataset enables a comprehensive analysis of socioeconomic and health factors, such as age, gender, geographic distribution, health-related behaviours, and their association with chronic illness in the Austrian population. These insights can guide and shape targeted public health and social policies across Europe.

Table 6 presents the distribution of chronic health conditions across a range of sociodemographic and lifestyle factors in a sample of 10,191 individuals. Participants with chronic conditions (n = 3,810) differed significantly across multiple variables compared to those without such conditions. Age showed a robust association with prolonged illness ($p < 0.001$), with affected participants having a mean age of 58.2 (±16.4) years, significantly older than the mean age of 47.1 (±17.7) years in the group without chronic conditions ($p < 0.001$). Gender also played a notable role ($p < 0.001$), with a higher proportion of females (55.1%) reporting chronic disorder compared to males (44.9%, $p < 0.001$).

Among geographic factors, federal states demonstrated significant variations in chronic illness prevalence ($p < 0.001$). For example, Vienna showed a higher proportion of individuals with chronic health conditions (22.8%) compared to other regions. Marital status was similarly significant ($p < 0.001$), with higher rates of chronic illness among widowed (12.1% with chronic conditions vs. 4.97% without it) and divorced individuals (13.1% with chronic conditions vs. 8.11% without it). In comparison, single participants reported lower rates (22.3% vs. 35.2%), indicating that marital status may be associated with chronic health issues.

Lifestyle factors, such as alcohol and smoking habits, were also significantly associated with chronic illness ($p < 0.001$). For instance, 29% of participants with these conditions reported abstaining from alcohol in the past 12 months, compared to 19.7% in the group without such disorders. Smoking frequency was similarly correlated with chronic illness ($p < 0.001$), with daily smoking rates slightly higher in the chronic illness group (19.2%) than in those without (18.9%). Physical activity frequency was another significant factor; those with chronic conditions were less

likely to engage in regular fitness activities ($p < 0.001$), with 32.2% of individuals in the chronic disorder group reporting no fitness activities, compared to 18.1% of those without chronic conditions. However, 13.2% of participants with chronic conditions engaged in heavy physical work, compared to 9.33% of those without, suggesting that heavy labour may contribute to the development of prolonged health disorders. Dietary habits also played an essential role in longstanding illness risk, with fruit and vegetable consumption showing significant associations. Those who rarely consumed vegetables ($p < 0.001$) or fruit ($p = 0.017$) had higher odds of persistent illness. Notably, 2.06% of individuals with chronic conditions reported never eating fruit, and 0.712% reported never eating vegetables.

Healthcare utilisation was significantly higher among individuals with chronic illnesses, with both general practitioner and specialist visit frequencies displaying clear trends ($p < 0.001$). For example, 21% of individuals with persistent conditions visited a general practitioner ten or more times in the past year, compared to only 3.08% of those without. This reflects the ongoing medical needs of individuals with longstanding disorders. The general state of health was another major factor, with those having chronic conditions less likely to report good or very good health ($p < 0.001$). Only 6.19% of individuals with chronic illness rated their health as "very good," compared to 44.6% of those without, while higher proportions in the chronic illness group rated their health as "poor" (15.2%) or "very poor" (4.37%). Restrictions in daily activities due to health problems were notably more common among individuals with chronic conditions ($p < 0.001$). Severe limitations affected 20.6% of those with persistent illness, compared to only 0.769% in the non-ill group, highlighting the substantial impact on daily functioning. Additionally, 66.8% of those with longstanding disorders reported experiencing activity restrictions lasting six months or more ($p < 0.001$), emphasising the enduring impact of these health issues. Utilisation of other medical services was also higher among those with chronic conditions ($p < 0.001$), with 63.8% of individuals with chronic conditions accessing additional services, compared to 33.8% of those without, further reflecting the complex healthcare needs of this group.

Employment status was another area of significant difference ($p < 0.001$). Individuals with chronic disorders were less likely to be employed or in apprenticeships (37.1%) compared to those without (61.3%). Prolonged illness rates were exceptionally high among retirees (50.1%) compared to those without longstanding conditions (24.9%), indicating a link between persistent illness and reduced workforce participation. Educational attainment was also inversely related to chronic disease ($p < 0.001$). Participants with only compulsory schooling were more likely to have chronic conditions (19.2% in the persistent illness group vs. 12.4% in those without). In comparison, university graduates had lower rates of chronic illness (15% in the prolonged illness group and 12.4% in those without). This pattern suggests that higher education may confer a protective effect against longstanding illness, possibly due to increased health literacy and better access to resources.

Finally, Body Mass Index (BMI) demonstrated a strong association with chronic illness ($p < 0.001$). Obesity was more prevalent among those with chronic conditions (23.3%) than those without (12.3%), indicating that a higher BMI might be a significant risk factor for chronic health issues.

Table 6: Distribution of chronic illness across sociodemographic variables and risk factors.

Variable [Missing]	Total (10,191)	Chronic illness (3,810)	No chronic disease (6,381)	*p*-value
Age (Mean/SD) [0]	51.2 (±18)	58.2 (±16.4)	47.1 (+-±17.7)	$p < 0.001$ (t-test)
Gender [0]				$p < 0.001$ (Chi-

Variable [Missing]	Total (10,191)	Chronic illness (3,810)	No chronic disease (6,381)	*p*-value
				square test)
- Female	5,392 (52.9%)	2,101 (55.1%)	3,291 (51.6%)	
- Male	4,799 (47.1%)	1,709 (44.9%)	3,090 (48.4%)	
Federal state [0]				p < 0.001 (Chi-square test)
- Burgenland	391 (3.84%)	158 (4.15%)	233 (3.65%)	
- Carinthia	732 (7.18%)	287 (7.53%)	445 (6.97%)	
- Lower Austria	1,972 (19.4%)	705 (18.5%)	1,267 (19.9%)	
- Salzburg	653 (6.41%)	190 (4.99%)	463 (7.26%)	
- Styria	1,501 (14.7%)	533 (14%)	968 (15.2%)	
- Tyrol	878 (8.62%)	327 (8.58%)	551 (8.64%)	
- Upper Austria	1,550 (15.2%)	587 (15.4%)	963 (15.1%)	
- Vienna	2,053 (20.1%)	867 (22.8%)	1,186 (18.6%)	
- Vorarlberg	461 (4.52%)	156 (4.09%)	305 (4.78%)	
Marital status [7]				p < 0.001 (Chi-square test)
- Divorced or dissolved registered partnership	1,015 (9.96%)	498 (13.1%)	517 (8.11%)	
- Married or in a registered partnership	5,298 (52%)	2,002 (52.6%)	3,296 (51.7%)	
- Single	3,094 (30.4%)	849 (22.3%)	2,245 (35.2%)	
- Widowed or surviving registered partner	777 (7.62%)	460 (12.1%)	317 (4.97%)	
Frequency of alcohol consumption in the last 12 months [366]				p < 0.001 (Chi-square test)
- Yes, daily	457 (4.48%)	223 (6.05%)	234 (3.81%)	
- Yes, a few times a week	2,198 (21.6%)	736 (20%)	1,462 (23.8%)	
- Yes, a few times a month	2,781 (27.3%)	854 (23.2%)	1,927 (31.4%)	
- Yes, a few times a year	2,110 (20.7%)	802 (21.8%)	1,308 (21.3%)	
- No, never	2,279 (22.4%)	1,070 (29%)	1,209 (19.7%)	
Frequency of smoking in the last 12 months [55]				p < 0.001 (Chi-

Variable [Missing]	Total (10,191)	Chronic illness (3,810)	No chronic disease (6,381)	*p*-value
				square test)
- Yes, daily	1,930 (18.9%)	729 (19.2%)	1,201 (18.9%)	
- Yes, a few times a week	270 (2.65%)	98 (2.58%)	172 (2.71%)	
- Yes, a few times a month	179 (1.76%)	49 (1.29%)	130 (2.05%)	
- Yes, a few times a year	212 (2.08%)	54 (1.42%)	158 (2.49%)	
- No, never	7,545 (74%)	2,863 (75.5%)	4,682 (73.8%)	
Municipality size classes [0]				p < 0.001 (Chi-square test)
- Vienna	2,053 (20.1%)	867 (22.8%)	1,186 (18.6%)	
- Other municipalities >100,000 pop.	837 (8.21%)	299 (7.85%)	538 (8.43%)	
- Municipalities >10,000 and <=100,000 pop.	1,552 (15.2%)	587 (15.4%)	965 (15.1%)	
- Municipalities <=10,000 pop.	5,749 (56.4%)	2,057 (54%)	3,692 (57.9%)	
Degree of urbanisation [0]				p < 0.001 (Chi-square test)
- Densely populated area	2,890 (28.4%)	1,166 (30.6%)	1,724 (27%)	
- Intermediate area	3,136 (30.8%)	1,159 (30.4%)	1,977 (31%)	
- Thinly populated area	4,165 (40.9%)	1,485 (39%)	2,680 (42%)	
OECD regional typology [3]				p = 0.012 (Chi-square test)
- Intermediate	2,644 (25.9%)	932 (24.5%)	1,712 (26.8%)	
- Predominantly Rural	4,378 (43%)	1,642 (43.1%)	2,736 (42.9%)	
- Predominantly Urban	3,166 (31.1%)	1,236 (32.4%)	1,930 (30.3%)	
Housing problem: air and water pollution, soot from traffic/industry [0]	762 (7.48%)	342 (8.98%)	420 (6.58%)	p < 0.001 (Chi-square test)
Physical activity level [5,007]				p < 0.001 (Chi-square test)
- Mostly sitting	2,506 (24.6%)	652 (47%)	1,854 (48.8%)	
- Mostly standing	699 (6.86%)	180 (13%)	519 (13.7%)	

Variable [Missing]	Total (10,191)	Chronic illness (3,810)	No chronic disease (6,381)	*p*-value
- Mostly walking or moderately strenuous physical activities	1,442 (14.1%)	373 (26.9%)	1,069 (28.2%)	
- Mostly heavy physical work	537 (5.27%)	183 (13.2%)	354 (9.33%)	
Frequency of fitness days [59]				p < 0.001 (Chi-square test)
- 2 times or more per day	359 (3.52%)	137 (3.61%)	222 (3.5%)	
- 1 time per day	1,426 (14%)	520 (13.7%)	906 (14.3%)	
- 4 to 6 times per week	1,600 (15.7%)	514 (13.6%)	1,086 (17.1%)	
- 1 to 3 times per week	3,400 (33.4%)	1,051 (27.7%)	2,349 (37%)	
- Less than 1 time per week	980 (9.62%)	348 (9.18%)	632 (9.97%)	
- Never	2,367 (23.2%)	1,221 (32.2%)	1,146 (18.1%)	
Consumption of fruit [52]				p = 0.017 (Chi-square test)
- 2 times or more per day	1,452 (14.2%)	554 (14.6%)	898 (14.2%)	
- 1 time per day	4,346 (42.6%)	1,597 (42.1%)	2,749 (43.3%)	
- 4 to 6 times per week	1,650 (16.2%)	599 (15.8%)	1,051 (16.6%)	
- 1 to 3 times per week	2,099 (20.6%)	783 (20.6%)	1,316 (20.7%)	
- Less than 1 time per week	402 (3.94%)	182 (4.8%)	220 (3.47%)	
- Never	190 (1.86%)	78 (2.06%)	112 (1.76%)	
Consumption of vegetables [56]				p < 0.001 (Chi-square test)
- 2 times or more per day	1,248 (12.2%)	434 (11.4%)	814 (12.8%)	
- 1 time per day	4,350 (42.7%)	1,569 (41.4%)	2,781 (43.8%)	
- 4 to 6 times per week	2,424 (23.8%)	909 (24%)	1,515 (23.9%)	
- 1 to 3 times per week	1,867 (18.3%)	767 (20.2%)	1,100 (17.3%)	
- Less than 1 time per week	183 (1.8%)	86 (2.27%)	97 (1.53%)	
- Never	63 (0.618%)	27 (0.712%)	36 (0.568%)	
Number of visits to the general practitioner in the last 12 months [27]				p < 0.001 (Chi-square test)

Variable [Missing]	Total (10,191)	Chronic illness (3,810)	No chronic disease (6,381)	*p*-value
- Never	1,715 (16.8%)	342 (8.99%)	1,373 (21.6%)	
- 1 to 2 times	3,768 (37%)	908 (23.9%)	2,860 (45%)	
- 3 to 5 times	2,813 (27.6%)	1,181 (31.1%)	1,632 (25.7%)	
- 6 to 9 times	874 (8.58%)	574 (15.1%)	300 (4.72%)	
- 10 times or more	994 (9.75%)	798 (21%)	196 (3.08%)	
Number of visits to a specialist in the last 12 months [23]				$p < 0.001$ (Chi-square test)
- Never	2,964 (29.1%)	660 (17.4%)	2,304 (36.2%)	
- 1 to 2 times	4,016 (39.4%)	1,222 (32.1%)	2,794 (43.9%)	
- 3 to 5 times	2,076 (20.4%)	1,081 (28.4%)	995 (15.6%)	
- 6 to 9 times	582 (5.71%)	420 (11%)	162 (2.54%)	
- 10 times or more	530 (5.2%)	419 (11%)	111 (1.74%)	
General state of health [47]				$p < 0.001$ (Chi-square test)
- Very good	3,065 (30.1%)	235 (6.19%)	2,830 (44.6%)	
- Good	4,058 (39.8%)	1,251 (33%)	2,807 (44.2%)	
- Moderate	2,216 (21.7%)	1,566 (41.3%)	650 (10.2%)	
- Poor	635 (6.23%)	578 (15.2%)	57 (0.898%)	
- Very poor	170 (1.67%)	166 (4.37%)	4 (0.063%)	
Restriction in daily activities due to health problems [8]				$p < 0.001$ (Chi-square test)
- Severely limited	834 (8.18%)	785 (20.6%)	49 (0.769%)	
- Somewhat limited	2,373 (23.3%)	1,872 (49.2%)	501 (7.86%)	
- Not limited	6,976 (68.5%)	1,151 (30.2%)	5,825 (91.4%)	
Restrictions due to health problems for at least six months [0]	2,986 (29.3%)	2,544 (66.8%)	442 (6.93%)	$p < 0.001$ (Chi-square test)
Utilization of necessary other medical services last 12 months [0]	4,586 (45%)	2,432 (63.8%)	2,154 (33.8%)	$p < 0.001$ (Chi-square test)
Current main activity [0]				$p < 0.001$ (Chi-

Variable [Missing]	Total (10,191)	Chronic illness (3,810)	No chronic disease (6,381)	*p*-value
				square test)
- Employed or apprentice	5,326 (52.3%)	1,412 (37.1%)	3,914 (61.3%)	
- In internship or civilian service	26 (0.255%)	3 (0.079%)	23 (0.36%)	
- In training	483 (4.74%)	86 (2.26%)	397 (6.22%)	
- Job seeker, unemployed	342 (3.36%)	184 (4.83%)	158 (2.48%)	
- Other	192 (1.88%)	45 (1.18%)	147 (2.3%)	
- Permanently unable to work for health reasons	82 (0.805%)	77 (2.02%)	5 (0.078%)	
- Retired	3,494 (34.3%)	1,907 (50.1%)	1,587 (24.9%)	
- Running a household	246 (2.41%)	96 (2.52%)	150 (2.35%)	
More than one job [0]	454 (4.45%)	124 (3.25%)	330 (5.17%)	$p < 0.001$ (Chi-square test)
Occupational function [4,450]				$p = 0.001$ (Chi-square test)
- Apprentice	106 (1.04%)	15 (0.966%)	91 (2.17%)	
- Civil servant	263 (2.58%)	82 (5.28%)	181 (4.32%)	
- Contractual employee	364 (3.57%)	113 (7.28%)	251 (5.99%)	
- Independent contractor	15 (0.147%)	4 (0.258%)	11 (0.263%)	
- Laborer	1,070 (10.5%)	319 (20.5%)	751 (17.9%)	
- Salaried employee	3,327 (32.6%)	848 (54.6%)	2,479 (59.2%)	
- Self-employed	589 (5.78%)	170 (10.9%)	419 (10%)	
- Unpaid employee in family business	7 (0.069%)	2 (0.129%)	5 (0.119%)	
Highest educational qualification [5]				$p < 0.001$ (Chi-square test)
- Compulsory school	1,519 (14.9%)	731 (19.2%)	788 (12.4%)	
- Apprenticeship with vocational school	3,441 (33.8%)	1,388 (36.5%)	2,053 (32.2%)	
- Technical or commercial college	1,342 (13.2%)	518 (13.6%)	824 (12.9%)	
- Matura	1,764 (17.3%)	531 (14%)	1,233 (19.3%)	
- Degree from a university or college	1,909 (18.7%)	569 (15%)	1,340 (21%)	
- Other qualification after the Matura	211 (2.07%)	69 (1.81%)	142 (2.23%)	

Variable [Missing]	Total (10,191)	Chronic illness (3,810)	No chronic disease (6,381)	*p*-value
Body height (Mean/SD) [118]	171 (±9.23)	170 (±9.31)	172 (±9.1)	p < 0.001 (t-test)
Body weight (Mean/SD) [255]	76.1 (±16.5)	78 (±17.7)	74.9 (±15.7)	p < 0.001 (t-test)
Total household income: net level [9,639]				p = 0.446 (Chi-square test)
- Up to 600 euros	4 (0.039%)	2 (0.971%)	2 (0.578%)	
- Up to 900 euros	11 (0.108%)	5 (2.43%)	6 (1.73%)	
- Up to 1,200 euros	20 (0.196%)	10 (4.85%)	10 (2.89%)	
- Up to 1,500 euros	21 (0.206%)	9 (4.37%)	12 (3.47%)	
- Up to 1,800 euros	27 (0.265%)	14 (6.8%)	13 (3.76%)	
- Up to 2,200 euros	63 (0.618%)	28 (13.6%)	35 (10.1%)	
- Up to 2,600 euros	52 (0.51%)	23 (11.2%)	29 (8.38%)	
- Up to 3,000 euros	81 (0.795%)	25 (12.1%)	56 (16.2%)	
- Up to 3,500 euros	39 (0.383%)	13 (6.31%)	26 (7.51%)	
- Up to 4,000 euros	74 (0.726%)	28 (13.6%)	46 (13.3%)	
- Up to 4,500 euros	29 (0.285%)	9 (4.37%)	20 (5.78%)	
- Up to 5,000 euros	60 (0.589%)	15 (7.28%)	45 (13%)	
- Up to 6,000 euros	35 (0.343%)	14 (6.8%)	21 (6.07%)	
- Up to 8,000 euros	22 (0.216%)	7 (3.4%)	15 (4.34%)	
- More than 8,000 euros	14 (0.137%)	4 (1.94%)	10 (2.89%)	
Body Mass Index (BMI) [262]				p < 0.001 (Chi-square test)
- Underweight (BMI value of less than 18.49)	249 (2.44%)	85 (2.28%)	164 (2.64%)	
- Normal (BMI value of 18.5 to 24.9)	4,581 (45%)	1,426 (38.3%)	3,155 (50.8%)	
- Overweight (BMI value of 25.0 to 29.9)	3,471 (34.1%)	1,343 (36.1%)	2,128 (34.3%)	
- Obese (BMI value of 30 or larger)	1,628 (16%)	866 (23.3%)	762 (12.3%)	

Figure 5 shows four bivariate maps illustrating the prevalence of chronic illness across Austrian regions in relation to tobacco use, alcohol intake frequency, average BMI, and average age. The top left map focuses on tobacco use across Austrian regions. Vienna displays darker shading, demonstrating higher rates of smoking, which is consistent with elevated chronic illness prevalence in this area. Conversely, regions like Salzburg and Stirya present lower tobacco use and a

corresponding reduction in chronic illness rates, supporting the hypothesis that lower smoking rates may protect against chronic disorders. The top right map highlights alcohol intake frequency (three times a week or more) and its association with longstanding illness. Upper Austria exhibits more frequent alcohol intake, which aligns with higher chronic illness prevalence in the area.

Furthermore, the bottom left map illustrates the average BMI across the regions where Burgenland, Upper, and Lower Austria have higher average BMI and, consequently, a higher prevalence of chronic illness. This trend highlights the potential drawback of high BMI for increasing chronic illness. Lastly, the bottom right map examines the average age in relation to chronic illness prevalence. Burgenland and Carinthia appear in darker shades, indicating an older population with a greater prevalence of chronic conditions. That aligns with the established association between ageing and increased susceptibility to prolonged health issues.

Figure 5: Discrepancies in chronic illness across Austrian regions by tobacco use, alcohol intake, BMI, and age.

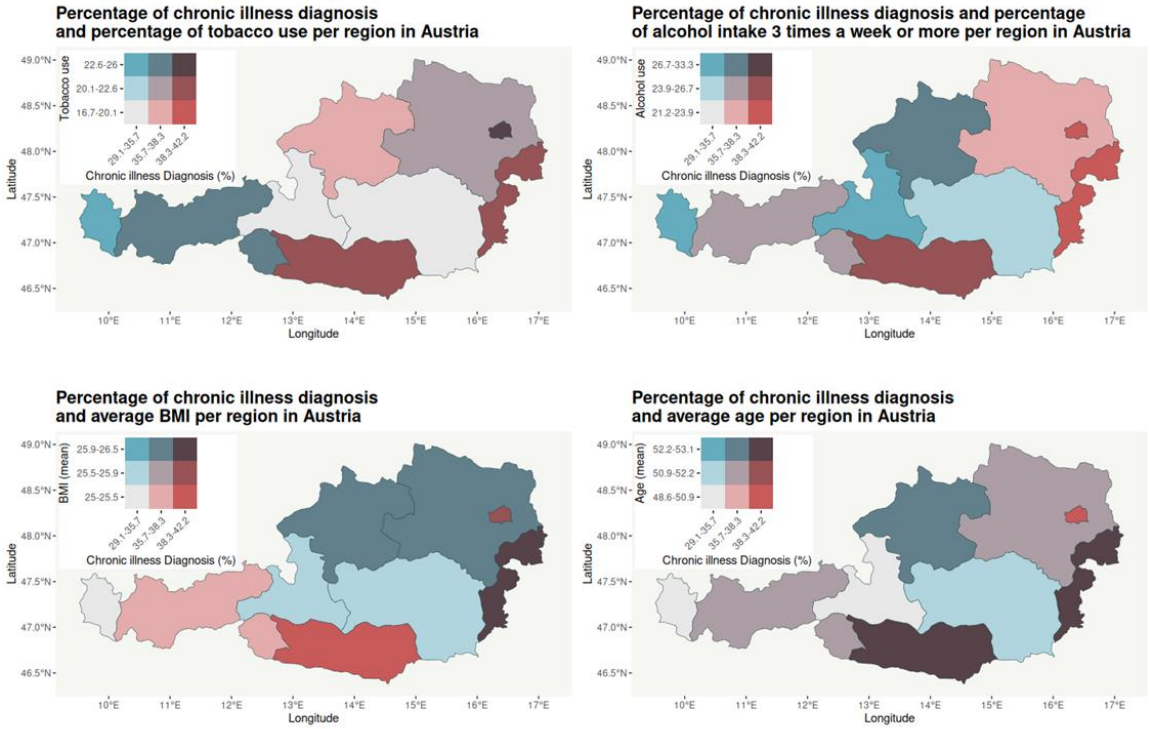


Table 7 presents predictors of chronic illness through logistic regression, focusing on age, gender, regional, and lifestyle variables. Age emerged as a robust predictor of chronic disease, with individuals aged 53-80 years showing significantly higher odds of having a chronic condition compared to younger participants (OR: 3.02, 95% CI: 2.78-3.29, $p < 0.001$). This finding is consistent with existing research linking ageing with increased susceptibility to chronic health issues, likely due to cumulative exposure to risk factors and the physiological decline associated with ageing. Gender also played a significant role, with males showing lower odds of reporting a chronic disorder than females (OR: 0.866, 95% CI: 0.799-0.939, $p < 0.001$), indicating a higher prevalence among females.

Regional differences were evident in the data, with participants residing in Salzburg showing lower odds of having a chronic condition compared to those in Burgenland (OR: 0.605, 95% CI: 0.465-

0.787, p < 0.001). Additionally, other municipalities with populations over 100,000 had reduced odds of chronic conditions compared to Vienna (OR: 0.76, 95% CI: 0.643-0.897, p = 0.001). These geographic variations suggest potential disparities in health determinants, such as healthcare access, environmental factors, and lifestyle influences that may impact chronic illness risk across regions.

Health behaviours had a notable influence on the likelihood of chronic illness. Daily alcohol consumption was associated with higher odds of having a chronic condition compared to less frequent drinking, such as a few times per month (OR: 0.465, 95% CI: 0.381-0.568, p < 0.001). A similar pattern was observed in smoking habits: individuals who smoked daily over the last 12 months presented higher odds of chronic conditions, whereas smoking only a few times per month (OR: 0.621, 95% CI: 0.438-0.868, p = 0.006) or a few times per week (OR: 0.563, 95% CI: 0.405-0.772, p < 0.001) was associated with a lower likelihood of chronic health issues.

Regarding physical activity levels, engaging primarily in heavy physical work was linked with higher odds of chronic conditions (OR: 1.47, 95% CI: 1.2-1.79, p < 0.001) in contrast to those with predominantly sedentary activities. Additionally, those who never engaged in fitness activities were at increased risk (OR: 1.73, 95% CI: 1.38-2.17, p < 0.001) compared to those exercising two or more times per week, highlighting the potential protective benefits of regular physical activity. Dietary habits further influenced chronic illness odds. Participants who consumed vegetables less frequently than twice a day had higher odds of developing longstanding conditions. For example, those who ate vegetables less than once a week had significantly increased odds (OR: 1.66, 95% CI: 1.21-2.27, p = 0.001). Similarly, individuals consuming fruit less than twice a day, such as those eating fruit less than once per week, also showed elevated odds of prolonged illnesses (OR: 1.34, 95% CI: 1.07-1.68, p = 0.01). These findings emphasise the importance of diet quality in preventing chronic disorders and support existing evidence linking a nutrient-poor diet to higher health risks.

Furthermore, healthcare utilisation frequency was a significant predictor of persistent illnesses. Participants who visited a general practitioner (GP) 10 times or more in the last year had markedly higher odds of having a chronic condition (OR: 16.3, 95% CI: 13.5-19.9, p < 0.001) in contrast to those with no GP visits. Likewise, frequent specialist visits were associated with a greater risk of chronic illness; individuals seeing a specialist ten or more times had 13.2 times higher odds of chronic disease than those with no specialist visits (OR: 13.2, 95% CI: 10.5-16.6, p < 0.001). These patterns likely reflect the increased medical needs and comorbidity management challenges of individuals with chronic conditions who require regular consultations.

Socioeconomic and demographic factors also played significant roles. Employment status (Current Main Activity) showed strong associations with chronic conditions (p < 0.001). Individuals who were retired had markedly higher odds of prolonged illness (OR: 3.33, 95% CI: 3.04-3.65, p < 0.001), as did those permanently unable to work due to health issues (OR: 42.7, 95% CI: 19.1-122, p < 0.001), compared to those employed or in apprenticeships. These findings underscore the health challenges faced by individuals no longer in the workforce, particularly those with disabilities or ongoing health restrictions. Educational attainment also influenced chronic illness odds; lower educational levels were linked to higher risk. Participants with only compulsory schooling had higher odds of chronic conditions compared to those with a university degree (OR: 0.458, 95% CI: 0.398-0.527, p < 0.001), suggesting that higher education may confer protective benefits through increased health literacy, healthier lifestyle choices, or improved access to healthcare resources.

Moreover, body weight was also a significant predictor, with participants weighing between 75 and 99 kg showing higher odds of developing a longstanding illness than those in the 34-75 kg range (OR: 1.49, 95% CI: 1.2-1.87, p < 0.001). Conversely, body height was inversely related to chronic

illness risk. Taller individuals (171 cm and above) had lower odds of reporting a persistent health condition compared to those shorter than 149 cm (OR: 0.652, 95% CI: 0.534-0.795, p < 0.001).

Table 7: Predictive factors for chronic illness, including sociodemographic and lifestyle variables.

Predictor	Chronic illness
Age: [16, 53]	1 [Referent]
Age: (53, 80]	3.02 (2.78, 3.29) [p < 0.001]
Gender: Female	1 [Referent]
Gender: Male	0.866 (0.799, 0.939) [p < 0.001]
Federal state: Burgenland	1 [Referent]
Federal state: Carinthia	0.951 (0.741, 1.22) [p = 0.695]
Federal state: Lower Austria	0.821 (0.658, 1.03) [p = 0.081]
Federal state: Salzburg	0.605 (0.465, 0.787) [p < 0.001]
Federal state: Styria	0.812 (0.647, 1.02) [p = 0.073]
Federal state: Tyrol	0.875 (0.686, 1.12) [p = 0.284]
Federal state: Upper Austria	0.899 (0.717, 1.13) [p = 0.357]
Federal state: Vienna	1.08 (0.866, 1.35) [p = 0.504]
Federal state: Vorarlberg	0.754 (0.57, 0.997) [p = 0.048]
Marital status: -2	1 [Referent]
Marital status: Divorced or dissolved registered partnership	5.78 (0.983, 109) [p = 0.105]
Marital status: Married or in a registered partnership	3.64 (0.622, 68.9) [p = 0.231]
Marital status: Single	2.27 (0.387, 42.9) [p = 0.448]
Marital status: Widowed or surviving registered partner	8.71 (1.48, 165) [p = 0.046]
Frequency of alcohol consumption in the last 12 months: Yes, daily	1 [Referent]
Frequency of alcohol consumption in the last 12 months: Yes, a few times a week	0.528 (0.431, 0.648) [p < 0.001]
Frequency of alcohol consumption in the last 12 months: Yes, a few times a month	0.465 (0.381, 0.568) [p < 0.001]
Frequency of alcohol consumption in the last 12 months: Yes, a few times a year	0.643 (0.525, 0.789) [p < 0.001]
Frequency of alcohol consumption in the last 12 months: No, never	0.929 (0.76, 1.14) [p = 0.471]
Frequency of smoking in the last 12 months: Yes, daily	1 [Referent]
Frequency of smoking in the last 12 months: Yes, a few times a week	0.939 (0.719, 1.22) [p = 0.639]
Frequency of smoking in the last 12 months: Yes, a few times a month	0.621 (0.438, 0.868) [p = 0.006]
Frequency of smoking in the last 12 months: Yes, a few times a year	0.563 (0.405, 0.772) [p < 0.001]
Frequency of smoking in the last 12 months: No, never	1.01 (0.909, 1.12) [p = 0.888]
Municipality size classes: Vienna	1 [Referent]

Predictor	Chronic illness
Municipality size classes: Other municipalities >100,000 pop.	0.76 (0.643, 0.897) [p = 0.001]
Municipality size classes: Municipalities >10,000 and <=100,000 pop.	0.832 (0.727, 0.952) [p = 0.008]
Municipality size classes: Municipalities <=10,000 pop.	0.762 (0.688, 0.845) [p < 0.001]
OECD regional typology: Intermediate	1 [Referent]
OECD regional typology: Predominantly Rural	1.1 (0.997, 1.22) [p = 0.057]
OECD regional typology: Predominantly Urban	1.18 (1.06, 1.31) [p = 0.003]
Physical activity level: Mostly sitting	1 [Referent]
Physical activity level: Mostly standing	0.986 (0.813, 1.19) [p = 0.887]
Physical activity level: Mostly walking or moderately strenuous physical activities	0.992 (0.855, 1.15) [p = 0.917]
Physical activity level: Mostly heavy physical work	1.47 (1.2, 1.79) [p < 0.001]
Frequency of fitness days: 2 times or more per day	1 [Referent]
Frequency of fitness days: 1 time per day	0.93 (0.733, 1.18) [p = 0.552]
Frequency of fitness days: 4 to 6 times per week	0.767 (0.606, 0.974) [p = 0.028]
Frequency of fitness days: 1 to 3 times per week	0.725 (0.58, 0.91) [p = 0.005]
Frequency of fitness days: Less than 1 time per week	0.892 (0.696, 1.15) [p = 0.371]
Frequency of fitness days: Never	1.73 (1.38, 2.17) [p < 0.001]
Consumption of fruit: 2 times or more per day	1 [Referent]
Consumption of fruit: 1 time per day	0.942 (0.833, 1.06) [p = 0.336]
Consumption of fruit: 4 to 6 times per week	0.924 (0.798, 1.07) [p = 0.287]
Consumption of fruit: 1 to 3 times per week	0.964 (0.84, 1.11) [p = 0.607]
Consumption of fruit: Less than 1 time per week	1.34 (1.07, 1.68) [p = 0.01]
Consumption of fruit: Never	1.13 (0.828, 1.53) [p = 0.44]
Consumption of vegetables: 2 times or more per day	1 [Referent]
Consumption of vegetables: 1 time per day	1.06 (0.928, 1.21) [p = 0.401]
Consumption of vegetables: 4 to 6 times per week	1.13 (0.976, 1.3) [p = 0.105]
Consumption of vegetables: 1 to 3 times per week	1.31 (1.13, 1.52) [p < 0.001]
Consumption of vegetables: Less than 1 time per week	1.66 (1.21, 2.27) [p = 0.001]
Consumption of vegetables: Never	1.41 (0.836, 2.34) [p = 0.192]
Number of visits to the general practitioner in the last 12 months: Never	1 [Referent]
Number of visits to the general practitioner in the last 12 months: 1 to 2 times	1.27 (1.11, 1.47) [p < 0.001]
Number of visits to the general practitioner in the last 12 months: 3 to 5 times	2.91 (2.53, 3.35) [p < 0.001]

Predictor	Chronic illness
Number of visits to the general practitioner in the last 12 months: 6 to 9 times	7.68 (6.4, 9.24) [p < 0.001]
Number of visits to the general practitioner in the last 12 months: 10 times or more	16.3 (13.5, 19.9) [p < 0.001]
Number of visits to a specialist in the last 12 months: Never	1 [Referent]
Number of visits to a specialist in the last 12 months: 1 to 2 times	1.53 (1.37, 1.7) [p < 0.001]
Number of visits to a specialist in the last 12 months: 3 to 5 times	3.79 (3.36, 4.29) [p < 0.001]
Number of visits to a specialist in the last 12 months: 6 to 9 times	9.05 (7.42, 11.1) [p < 0.001]
Number of visits to a specialist in the last 12 months: 10 times or more	13.2 (10.5, 16.6) [p < 0.001]
The general state of health: Very good	1 [Referent]
The general state of health: Good	5.37 (4.63, 6.24) [p < 0.001]
General state of health: Moderate	29 (24.7, 34.2) [p < 0.001]
The general state of health: Poor	122 (91, 167) [p < 0.001]
The general state of health: Very poor	500 (209, 1,634) [p < 0.001]
Restriction in daily activities due to health problems: Severely limited	1 [Referent]
Restriction in daily activities due to health problems: Somewhat limited	0.233 (0.17, 0.313) [p < 0.001]
Restriction in daily activities due to health problems: Not limited	0.012 (0.009, 0.016) [p < 0.001]
Utilization of necessary other medical services last 12 months: FALSE	1 [Referent]
Utilization of necessary other medical services last 12 months: TRUE	3.46 (3.18, 3.77) [p < 0.001]
Current main activity: Employed or apprentice	1 [Referent]
Current main activity: In internship or civilian service	0.362 (0.086, 1.04) [p = 0.098]
Current main activity: In training	0.6 (0.469, 0.76) [p < 0.001]
Current main activity: Job seeker, unemployed	3.23 (2.59, 4.03) [p < 0.001]
Current main activity: Other	0.849 (0.598, 1.18) [p = 0.343]
Current main activity: Permanently unable to work for health reasons	42.7 (19.1, 122) [p < 0.001]
Current main activity: Retired	3.33 (3.04, 3.65) [p < 0.001]
Current main activity: Running a household	1.77 (1.36, 2.3) [p < 0.001]
Occupational function: Apprentice	1 [Referent]
Occupational function: Civil servant	2.75 (1.54, 5.2) [p = 0.001]
Occupational function: Contractual employee	2.73 (1.56, 5.1) [p < 0.001]
Occupational function: Independent contractor	2.21 (0.556, 7.45) [p = 0.221]
Occupational function: Laborer	2.58 (1.51, 4.69) [p < 0.001]
Occupational function: Salaried employee	2.08 (1.23, 3.75) [p = 0.009]
Occupational function: Self-employed	2.46 (1.43, 4.53) [p = 0.002]

Predictor	Chronic illness
Occupational function: Unpaid employee in family business	2.43 (0.327, 12.4) [p = 0.315]
Highest educational qualification: Compulsory school	1 [Referent]
Highest educational qualification: Apprenticeship with vocational school	0.729 (0.645, 0.823) [p < 0.001]
Highest educational qualification: Technical or commercial college	0.678 (0.584, 0.786) [p < 0.001]
Highest educational qualification: Matura	0.464 (0.402, 0.536) [p < 0.001]
Highest educational qualification: Degree from a university or college	0.458 (0.398, 0.527) [p < 0.001]
Highest educational qualification: Other qualification after the Matura	0.524 (0.384, 0.708) [p < 0.001]
Body height: [138, 170]	1 [Referent]
Body height: (170, 210]	0.718 (0.662, 0.779) [p < 0.001]
Body weight: [34, 75]	1 [Referent]
Body weight: (75, 190]	1.28 (1.18, 1.39) [p < 0.001]

3.1.4 European Prospective Investigation into Cancer and Nutrition (EPIC)

EPIC is a large-scale cohort study that examines how dietary habits, lifestyle factors, and environmental exposures contribute to cancer and other chronic disease risks. Initiated in the 1990s, EPIC involves more than 500,000 participants across 10 European countries, making it one of the most extensive studies in nutritional epidemiology. This prospective study collects baseline information on diet, physical activity, and lifestyle factors from participants before disease development, which allows researchers to track disease outcomes over time and study associations between exposures and health risks.[15,16]

In addition to baseline data, EPIC has collected biospecimens like blood samples, enabling biomarker research to deepen understanding of how different biological mechanisms are influenced by diet and lifestyle. The study’s breadth, including its large participant base and geographic diversity, provides comprehensive insights into the links between cancer, diet, and lifestyle factors in different populations. EPIC’s findings have informed numerous public health recommendations, supporting initiatives that address cancer prevention through dietary and lifestyle adjustments.

A proposal has been submitted to gain access to the EPIC dataset. EPIC is a significant resource for research on the relationships between dietary, lifestyle, and environmental factors and the incidence of cancer. Access to this dataset will facilitate the identification of behavioural and psychosocial risk factors that influence health outcomes and cancer survivorship across EU member states. The outcome of the proposal is awaited, and it will determine the possibility of utilising this dataset in the iBeChange project.

3.2 Aggregate data

Beyond the three individual-level databases described above, we also performed detailed analyses on two of the 25 databases with aggregate data. Table 8 presents the list of the 25 databases with aggregate data to which we obtained access. Below, we describe the European Health Interview Survey (EHIS) and the European Health for All database (HFA-DB) in more detail, as these

databases provided comprehensive information on key health indicators and risk factors relevant to our study. EHIS offered extensive data on health behaviours, lifestyle factors, and chronic conditions across European populations, allowing us to assess patterns in health behaviours associated with cancer risk. Meanwhile, the HFA-DB provided valuable country-level statistics on healthcare access, socioeconomic conditions, and environmental factors, which enabled us to contextualise our findings within broader public health trends across Europe.

Table 8: Downloaded databases with aggregate data.

Database	Description and variables
Global Burden of Disease Study 2019 (GBD 2019) Risk Factors-Attributable Cancer Burden Estimates 2010–2019	Provides comprehensive data on cancer burden attributed to multiple modifiable risk factors across the period from 2010 to 2019. This dataset covers numerous physical and lifestyle-related factors, including physical activity, tobacco use, alcohol consumption, dietary habits, and high BMI, across various types of cancer. However, it excludes psychosocial variables, maintaining a focus on measurable physical health and lifestyle factors. Available at the country level, it includes data worldwide, enabling users to analyse and compare cancer risk across different regions and Socio-demographic Index (SDI) classifications. This global dataset allows for detailed insights into cancer prevention, offering essential information for public health planning.
European Health Interview Survey (EHIS)	The European Health Interview Survey (EHIS) from 2019 is an interviewer-administered questionnaire conducted across EU member states and coordinated by Eurostat. It aims to provide reliable and comparable data on health status, healthcare utilisation, and health behaviour determinants among residents aged 15 and older. This dataset is essential for monitoring health trends and evaluating public health policies, particularly concerning psychosocial and behaviour-related variables such as smoking, alcohol consumption, diet, physical activity, and chronic conditions, including cancer.
OECD (Organisation for Economic Co-operation and Development)	The Organisation for Economic Co-operation and Development (OECD) is a significant international organisation focused on enhancing policies to improve lives. It provides health-related datasets, including statistics on cancer incidence rates in various EU member countries such as Austria, Belgium, Denmark, France, and Germany, covering the period from 2000 to 2012.
Indicadores clave del Sistema Nacional de Salud (España)	The Key Indicators of the National Health System (INCLASNS) in Spain offer a concise set of metrics that are essential for evaluating the performance of the health system. The database includes several variables such as age, gender, prevalence, and specific health behaviours, including fruit and vegetable consumption, alcohol intake, tobacco use, and obesity rates for the years 2017 and 2020. However, it does not provide specific data on cancer.
NIJZ data portal (Slovenia)	The National Institute of Public Health (NIJZ) in Slovenia is dedicated to studying, protecting, and enhancing the population's health through public awareness and preventive measures. Authorised to collect health and healthcare data, NIJZ provides valuable datasets that encompass multiple variables, including alcohol consumption, bodily pain, body mass index (BMI), nutrition, physical activity, preventive services, self-assessment of health and chronic diseases, depression symptoms, tobacco use, educational attainment, employment status, perceived social support, and social network size. The datasets cover the years 2014 and 2019, offering

	insights into health determinants and outcomes in the Slovenian population. However, it does not provide specific data about cancer.
Monitoring system - Study of habits affecting the health of the population of Latvia (Latvia)	The Center for Disease Prevention and Control (SPKC) in Latvia operates under the authority of the Ministry of Health, focusing on implementing public health policy related to epidemiological safety and disease prevention alongside healthcare quality. The Monitoring System includes various datasets that track health habits affecting the Latvian population, covering aspects such as cancer incidence, mortality rates, and health behaviours from 2008 to 2022. This data allows for a detailed analysis of public health trends, enabling researchers and policymakers to assess the effectiveness of health promotion strategies, monitor risk factors such as smoking and alcohol consumption, and evaluate the overall health status of different demographics.
Slovenia and cancer - SLORA	The Slovenia Cancer Registry, managed by the Oncological Institute of Ljubljana, has been operational since 1950 and is among the oldest population-based cancer registries in Europe. It provides comprehensive epidemiological data on cancer incidence, mortality, prevalence, and survival, with records available from 1961 onwards for some variables. The registry includes information on various cancers, helping assess the impact of primary and secondary prevention strategies and informing public health policies. However, it does not provide psychosocial data.
Inserm CépiDc (France)	Established in 1964, Inserm (Institut National de la Santé et de la Recherche Médicale) is a prominent public scientific and technological institution in France, operating under the joint supervision of the Ministry of Health and the Ministry of Research. The Inserm CépiDc dataset provides mortality statistics related to cancer in France, covering the years 1979 to 2021. It includes detailed information on causes of death for various malignant tumours, including colorectal, lung, and breast cancers, categorised by sex and age group, but it lacks psychosocial or behavioural-related data on the French population.
StatFin database (Finland)	The StatFin database offers a comprehensive collection of statistical tables reflecting Finnish society and its demographics. The database provides free access to a wide range of statistical data, including detailed mortality statistics related to cancer from 1971 to 2022. It encompasses various cancer types, including malignant neoplasms of the colon, rectum, lung, and breast, and allows users to analyse mortality data by age, sex, and underlying cause of death. Yet, it does not provide psychosocial data.
National Board of Health and Welfare's Cancer Register (Sweden Cancer Registry)	The National Board of Health and Welfare's Cancer Register (NKR) is the authoritative source for cancer statistics in Sweden, providing comprehensive data on malignant tumours diagnosed within the country. It captures around 60,000 new cancer cases annually, including essential details such as gender, age, residence at diagnosis, and clinical information. The NKR collects data on incidence, mortality, and survival rates, making it crucial for research and public health initiatives. Data on the date of diagnosis, metastases, and death, as well as migration status, are also integrated, ensuring a robust framework for monitoring cancer trends and evaluating healthcare outcomes. However, the database lacks information on psychosocial variables.
Instituto Nacional de Estadística (INE, Spain)	The Instituto Nacional de Estadística (INE) is an autonomous administrative body under the Ministry of Economy, Commerce, and Enterprise, established to oversee public statistical functions in Spain. The

	INE is tasked with conducting extensive statistical operations, including demographic and economic censuses, national accounts, and various social and economic indicators. The database includes vital mortality statistics detailing causes of death in Spain from 1980 to 2022, with a focus on multiple cancer types, including malignant tumours of the colon, rectum, lungs, and breast. The dataset provides insights by sex and age, enabling researchers and policymakers to analyse mortality trends and inform public health strategies effectively.
Netherlands Cancer Registry (NCR)	The Netherlands Cancer Registry (NKR) is a national database managed by the Integraal Kankercentrum Nederland (IKNL), providing reliable data on all cancer patients in the country. The NKR encompasses various metrics, including cancer incidence from 1989 to 2023, mortality data from 1989 to 2022, and prevalence figures from 2009 to 2023. It also includes detailed information on age, gender, cancer type, and survivors, facilitating a thorough understanding of cancer trends and patient outcomes in the Netherlands. However, the database lacks information on psychosocial variables.
Statistik Austria	Statistik Austria maintains the Austrian National Cancer Register, which serves as a comprehensive resource for cancer statistics in the country. Key indicators include cancer incidence, mortality, the prevalence of individuals living with a cancer diagnosis, and survival probabilities. The database also contains information on hospitalisation and social position, providing valuable context for understanding cancer's impact on different demographic groups. It employs the International Statistical Classification of Diseases and Related Health Problems (ICD-10) for tumour classification, ensuring standardised reporting and analysis of cancer data.
Statistics Poland	Statistics Poland has a high-value dataset with six thematic categories, including national health statistics related to cancer incidence and mortality. This database aggregates cancer data, enabling researchers and policymakers to analyse cancer trends and inform public health initiatives effectively. The database includes detailed statistics on various cancer types, stratified by sex, location of neoplasms, and total cases for the years 2019, 2020, and 2021. Yet, the database does not have information on psychosocial variables.
WHO mortality database	The WHO Mortality Database is a valuable resource for comparative epidemiological studies that focus on mortality causes, including various types of cancer. It offers extensive data collected from national vital registration systems spanning from 1950 to 2021. This database includes detailed statistics on causes of death, allowing users to examine mortality trends across different regions and demographics worldwide. It contains over 2 million records of deaths, all coded by national authorities, with each Member State reporting both its population and mortality figures. However, the database does not provide information on psychosocial variables.
Global Cancer Observatory: Cancer Over Time WHO	Cancer Over Time provides extensive international retrospective data on cancer-specific incidence and mortality rates from national and subnational population-based cancer registries. This data is part of the Cancer Incidence in Five Continents (CI5) CI5plus database and covers the years from 1992 to 2014. The database includes detailed statistics for 33 cancer sites, stratified by year, sex, and age group. The mortality data is sourced from the World Health Organization (WHO) Mortality

	Database. Cancer Over Time employs age-standardized rates to ensure comparability across different populations, while changes in coding and certification practices are acknowledged to potentially affect longitudinal data. However, it does not provide psychosocial variables.
Global Cancer Observatory: Cancer Today WHO	Cancer Today offers a genuine assessment of the global cancer burden in 2022, leveraging GLOBOCAN estimates for incidence, mortality, and prevalence across 185 countries and territories, covering 36 cancer types differentiated by sex and age group. The International Agency for Research on Cancer (IARC) collaborates with national staff to enhance local data quality and registry coverage through initiatives like the Global Initiative for Cancer Registry Development (GICR). Cancer Today provides insights into cancer incidence, mortality, and prevalence. The methodology used for estimation is country-specific and relies on data from population-based cancer registries and the World Health Organization. Yet, it does not cover psychosocial variables.
Global Cancer Observatory: Cancer Tomorrow WHO	Cancer Tomorrow is a forward-looking project that offers a suite of data visualisation tools to predict cancer incidence and mortality from current estimates in 2022 through 2050. Utilising data from the GLOBOCAN project, the platform provides projections for 36 specific cancer types and all cancer sites combined across 185 countries and territories, stratified by sex and age group. The predictions are based on the assumption that the national cancer rates observed in 2022 will remain stable throughout the prediction period, combined with accurate national population projections. However, it does not provide psychosocial variables.
Global Cancer Observatory: Cancer Causes WHO	The Global Cancer Observatory (GCO), managed by the International Agency for Research on Cancer (IARC), serves as a vital platform for disseminating and visualising cancer statistics globally. The Cancer Causes section provides estimates of cancer cases attributable to obesity and alcohol consumption in 2012 and 2021. The project highlights the proportion and number of preventable cancer cases linked to these factors, utilising body mass index (BMI) estimates and alcohol consumption data stratified by age and sex.
NORDCAN Data	NORDCAN is a comprehensive database of cancer statistics for the Nordic countries, including Denmark, Finland, Iceland, Norway, Sweden, the Faroe Islands, and Greenland. It provides detailed information on cancer incidence, mortality, prevalence, and survival rates. Data is sourced from national cancer registries and cause of death registries. The database is updated annually and covers more than 50 cancer types. It features advanced prediction models that forecast cancer incidence and mortality trends over the next 20 years based on historical data and population forecasts. However, it does not cover psychosocial variables.
ECIS - European Cancer Information System	The European Cancer Information System (ECIS) serves as an essential resource for the latest cancer burden indicators across Europe, facilitating the exploration of geographical patterns and temporal trends in cancer incidence, prevalence, mortality, and survival data for major cancer types. The database also includes 2040 incidence estimates for EU-27 countries derived from cancer registries participating in the IARC’s Cancer Incidence in Five Continents Volume XII. It offers survival estimates based on data from 99 cancer registries covering the period from 2000 to 2007. However, it does not cover psychosocial variables.

Cancer Incidence in Five Continents (CI5): WHO	CI5 represents a collaborative effort between the International Agency for Research on Cancer and the International Association of Cancer Registries, providing a comprehensive overview of cancer incidence patterns globally. Published every five years, this series compiles data from population-based cancer registries at both subnational and national levels, offering key insights into changing cancer trends from 1953 to 2012. The CI5 database variables include cancer site-specific incidence rates, age-standardized rates, and crude rates, facilitating the analyses by sex and age group. The dataset also provides information on cases of unknown age and at-risk populations, contributing to world-standardized incidence calculations. However, it does not include psychosocial variables.
European Health for All database (HFA-DB)	HFA-DB is a repository of health statistics for European countries, managed by the World Health Organization (WHO) Regional Office for Europe. It supports evidence-based public health policy by providing reliable and comparable data on health status, risk factors, and healthcare systems. The database encompasses a wide range of indicators related to morbidity, mortality, health determinants, and healthcare utilisation, enabling cross-national comparisons. HFA-DB is essential for monitoring public health trends and addressing health disparities, ultimately aiding efforts to improve health outcomes and promote health equity across Europe.
EUROpean CAncer REgistry based study on survival and care of cancer patients (EUROCARE)	EUROCARE is the largest collaborative research project focused on population-based cancer survival in Europe. The project aims to monitor cancer survival trends, measure cancer prevalence, and analyse patterns of care among cancer patients across European countries. It encompasses data from 117 cancer registries across 31 European countries, including approximately 28 million cancer diagnoses, supported by the European Network of Cancer Registries (ENCR). The dataset includes variables such as cancer site, patient demographics (age, gender), time and period from diagnosis, and survival indicators (e.g., observed, expected, and relative survival rates). However, it does not include psychosocial variables.
Lithuanian Cancer Registry (NVI)	The Lithuanian Cancer Registry is responsible for the official registration of oncological disease cases across Lithuania. The registry collects notifications regarding the first diagnosis of oncological diseases from personal healthcare institutions, processing over 20,000 new cases annually. The dataset includes various fields such as age group, patient county, gender, year of diagnosis, and various clinical and pathological classifications. Additionally, it captures cancer details on multiplicity, laterality, localisation (including ICD-10 codes), and morphology differentiation.

3.2.1 European Health Interview Survey (EHIS)

The European Health Interview Survey (EHIS) of 2019 is an interviewer-administered questionnaire conducted across European Union member states and coordinated by Eurostat, the statistical office of the European Union.[17] The EHIS aims to provide reliable, comparable data on the health status, healthcare use, and determinants of health behaviour among residents aged 15 and older. This dataset plays a critical role in monitoring health trends and evaluating the impact

of public health policies across Europe, particularly in areas such as smoking, alcohol consumption, diet, physical activity, and chronic health conditions, including cancer.

Cancer remains one of the leading causes of morbidity and mortality within the European Union,[18] prompting an urgent need to explore potential correlations between cancer rates and modifiable risk factors captured in the EHIS. Through the 2019 dataset, researchers and policymakers have the opportunity to identify associations between cancer mortality and behavioural factors. Understanding these relationships is crucial for developing evidence-based interventions and preventive strategies aimed at reducing cancer risk and improving public health outcomes across Europe.

Figure 6 presents three bivariate maps illustrating cancer mortality rates across European countries in different BMI categories: underweight, normal, and obese in 2019. The first map (top left) depicts the association between cancer mortality and the percentage of underweight individuals. Poland and Hungary exhibit the strongest association, indicating high cancer mortality alongside a significant proportion of underweight individuals. Conversely, countries in Northern Europe, like Norway and Finland, show lower cancer mortality rates and lower prevalence of underweight individuals.

The second map (top right) shows cancer mortality rates in relation to the prevalence of individuals with a normal BMI. Denmark displays the darkest shade, signifying a strong association where high cancer mortality is coupled with a higher proportion of individuals with normal BMI. This pattern may imply that a normal BMI does not necessarily mitigate cancer mortality in this region, potentially highlighting the role of additional risk factors in the development of cancer. In contrast, countries like Finland, Spain, and Iceland, with lighter shades, display lower associations, suggesting that nations with lower cancer mortality also have fewer individuals in the normal BMI category.

The third map (bottom) correlates cancer mortality with obesity prevalence, with several countries showing a strong interrelationship. Estonia, Latvia, Czech Republic, Slovakia, Hungary, Slovenia, and Croatia are on the list, indicating high cancer mortality alongside elevated obesity levels. That supports the established link between obesity and increased cancer risk described in the literature.[19] On the other hand, many other countries, such as Norway, Sweden, Spain, France, Belgium, and Italy, display lighter shades, reflecting lower obesity rates and lower cancer mortality, suggesting a weaker association and possibly underscoring the benefits of lower obesity prevalence in reducing cancer risk.

Figure 6: Cancer mortality rate and Body Mass Index (BMI) categories across European countries in 2019.

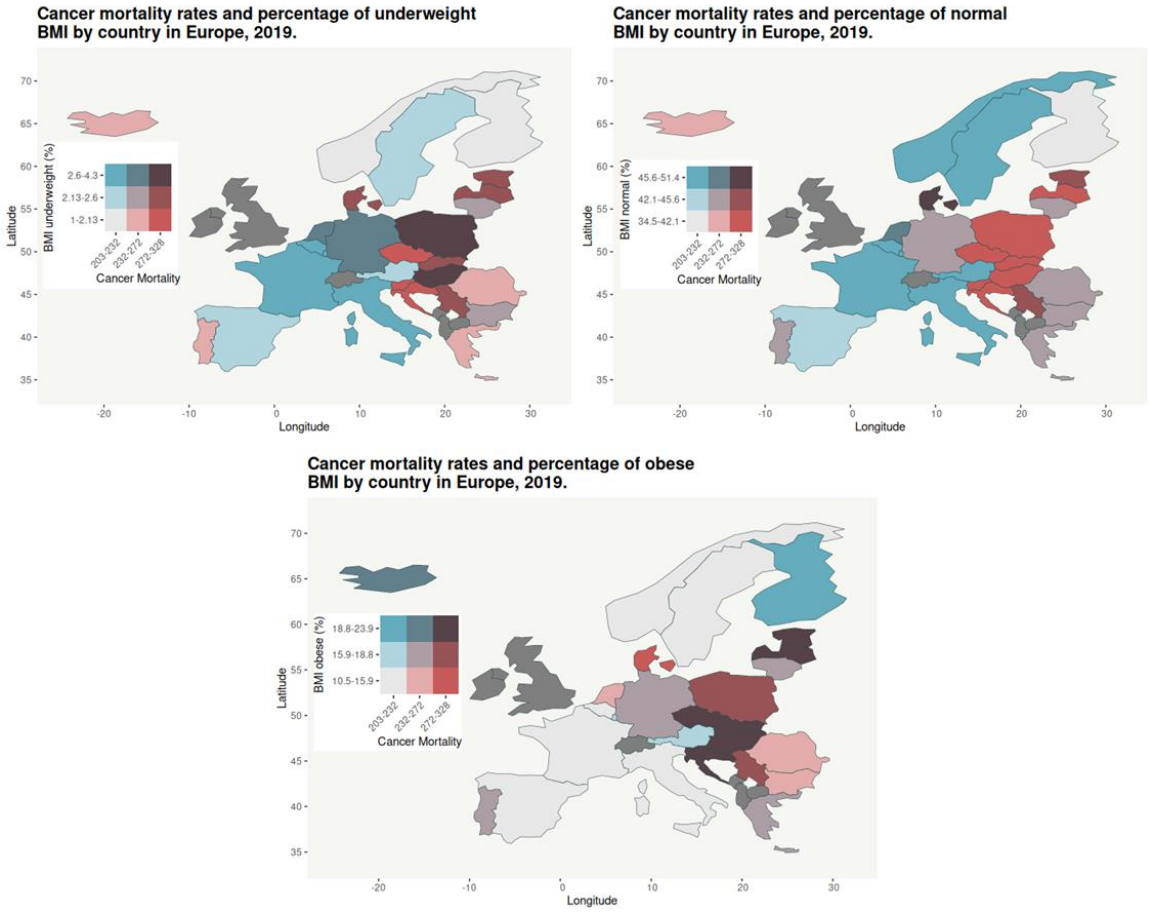


Figure 7 shows two bivariate maps illustrating the association between cancer mortality rates and alcohol consumption in the last 12 months across European countries in 2019. The map on the left correlates daily alcohol consumption with cancer mortality. Eastern European countries, along with some Southern nations like Croatia, show the darkest colours, reflecting a high percentage of daily alcohol consumption alongside elevated cancer mortality rates. Additionally, Denmark also presents the same pattern. This result may suggest a relationship between regular alcohol consumption and increased cancer risk. On the other hand, countries in Northern Europe, such as Norway and Sweden, exhibit lighter shades, reflecting lower daily alcohol consumption and lower cancer mortality rates.

The second map (right) examines cancer mortality rates relative to the prevalence of individuals who never drank or abstained from drinking in the past year. Again, Croatia exhibits the darkest pattern together with Hungary and Serbia, indicating that even among populations with a higher prevalence of non-drinkers or those abstaining from alcohol in the last 12 months, cancer mortality may remain high. This pattern may suggest the influence of other risk factors driving cancer mortality beyond alcohol consumption alone in these countries. Conversely, Norway and Sweden have lower cancer mortality and diminished rates of alcohol abstention.

Figure 7: Cancer mortality rate and frequency of alcohol intake across European countries in 2019.

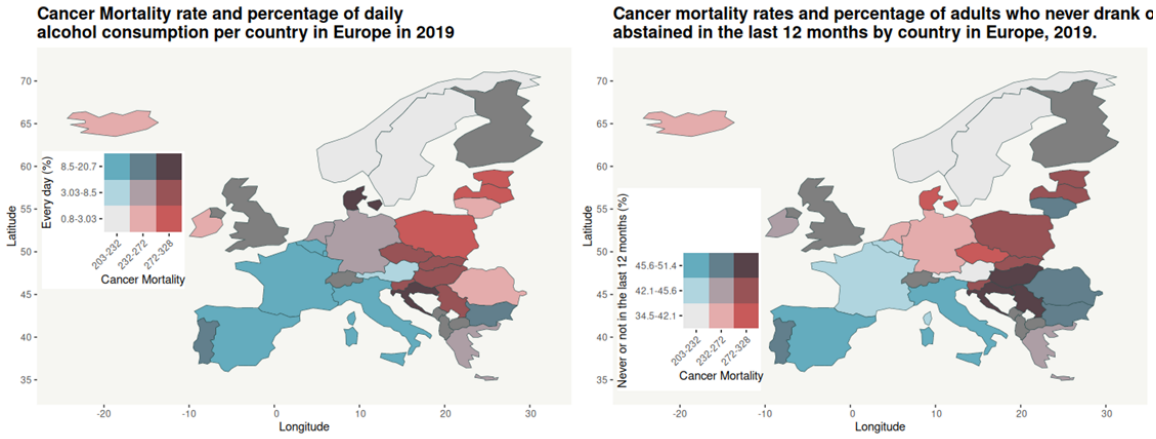


Figure 8 reveals insights into the link between smoking habits and cancer mortality across Europe. On the left map, the percentage of non-smokers alongside cancer mortality rates can be found. Countries like Serbia, Hungary, Slovakia, and the Czech Republic expose a high cancer mortality rate within low non-smoking populations, indicating that smoking may be driving the high cancer burden in Eastern Europe. However, Denmark shows a high percentage of non-smokers yet still has a high cancer mortality rate, suggesting that other factors may contribute to cancer incidence in this region. Meanwhile, Western and Southern European countries like France and Spain appear to show both lower cancer mortality and a relatively low percentage of non-smokers. This pattern suggests that while these regions have a high proportion of smokers, other factors may be mitigating the high cancer burden.

Shifting the focus to the right map, which explores daily smoking prevalence against cancer mortality. Overall, we observe a starkly different picture with the darkest colours around the continent. In Slovakia, Hungary, Serbia, Croatia, and Latvia, high smoking rates align closely with high cancer mortality. This alignment reinforces the well-established link between smoking and increased cancer risk,[20] painting a clear picture of where public health interventions could make the most impact. On the other hand, Nordic countries like Norway, Sweden, and Finland appear in lighter shades, where both daily smoking rates and cancer mortality are lower, underscoring the potential protective effect of lower smoking prevalence on cancer outcomes.

Figure 8: Cancer mortality rate and smoking across European countries in 2019.

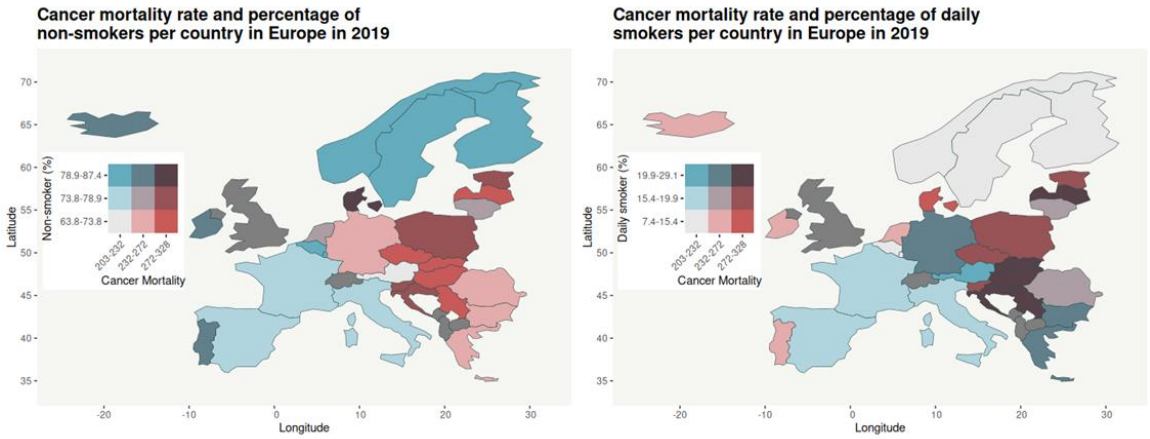


Figure 9 examines the relationship between cancer mortality and the daily consumption of fruits and vegetables across Europe. The left map shows that in most countries where a high percentage of the population does not consume fruits or vegetables daily, the cancer mortality rate is high. For instance, Eastern European nations, such as the Czech Republic and Slovakia, together with Latvia and Serbia, show the strongest association. This suggests that insufficient consumption of fruits and vegetables may contribute to elevated cancer mortality in these regions, aligning with the understanding that a poor diet can be a significant cancer risk factor. In contrast, Norway, Spain, France, and Italy appear in lighter shades, with both lower cancer mortality and fewer people reporting zero fruit and vegetable consumption, hinting at a possible protective effect of a better diet.

On the right map, depicting cancer mortality over consuming five or more portions of fruits and vegetables daily, an expected pattern emerges. Eastern European countries again show darker shades, where high cancer mortality persists together with a low percentage of the population eating five or more portions of fruits and vegetables daily, reinforcing that eating fewer fruits and vegetables may significantly higher the chances of dying of cancer. Meanwhile, nations with higher fruit and vegetable consumption, such as Finland, France, Belgium, and Luxemburg, reveal lower cancer mortality, supporting the idea that consistent intake of this food group could play a role in reducing cancer mortality.

Figure 9: Cancer mortality rate and daily consumption of fruit and vegetables across European countries in 2019.

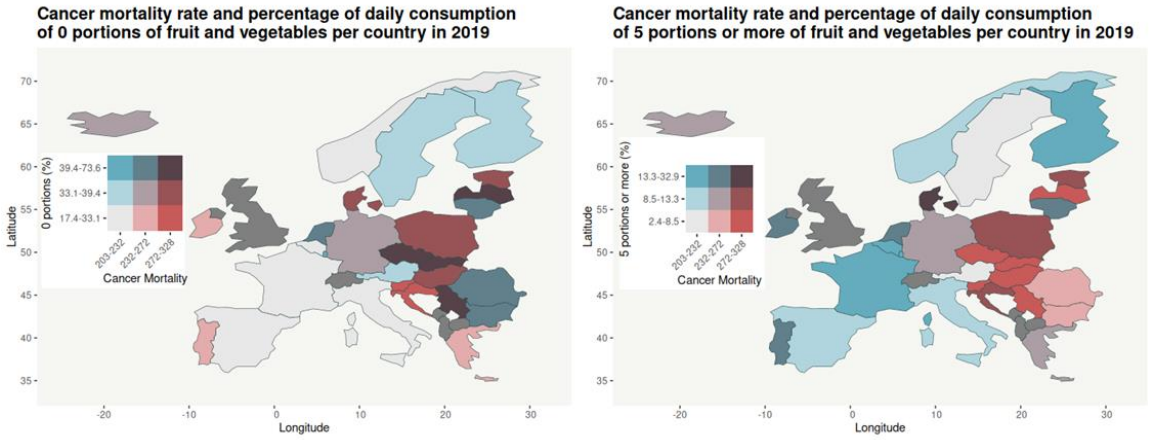
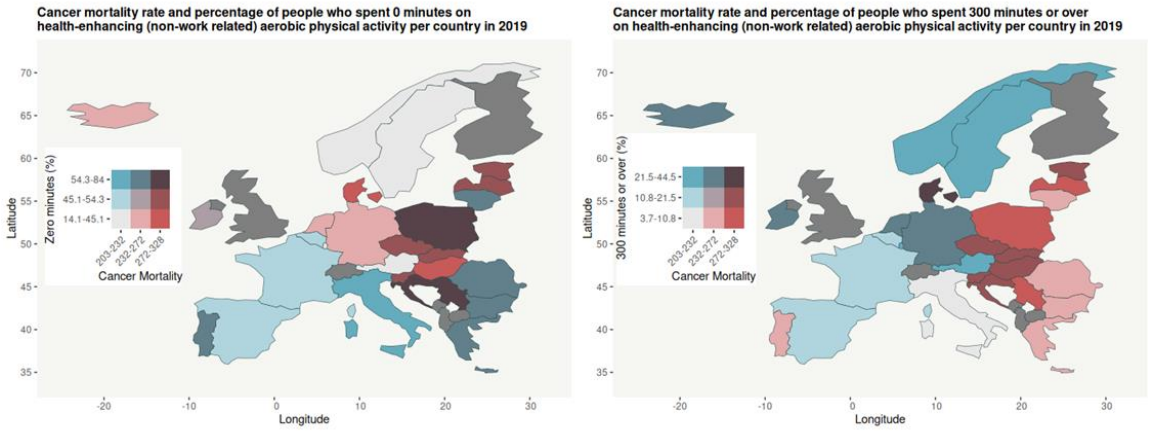


Figure 10 presents the association between cancer mortality rates and time spent on health-enhancing (non-work-related) aerobic physical activity across Europe in 2019. On the left map, Romania, Poland, Croatia, and Serbia appear in dark shades, reflecting high cancer mortality rates paired with a significant portion of the population reporting zero minutes of health-enhancing physical activity. This pattern highlights the potential impact of physical inactivity on cancer mortality, suggesting that sedentary lifestyles in these regions may contribute to the elevated cancer burden. On the other hand, Spain, France, Belgium, and most importantly, Luxemburg, Norway, and Sweden show lighter shades, indicating both lower cancer mortality and a smaller percentage of inactive individuals, supporting the idea that higher physical activity levels may mitigate cancer risk.

In contrast, the right map reveals the relationship between cancer mortality and individuals achieving 300 minutes or more of aerobic physical activity weekly. Countries like Germany and The Netherlands display relatively high cancer mortality despite higher activity levels, suggesting that factors beyond physical activity may influence cancer outcomes in these regions. On the opposite, an expected pattern is seen in Serbia, Poland, and Latvia, where high cancer mortality rates are accompanied by a low percentage of people practising physical activities for 300 minutes or over in 2019. Conversely, Nordic countries, including Sweden and Norway, show an ideal scenario with lower cancer mortality and higher physical activity levels, reinforcing the role of regular exercise in reducing cancer risk.

Figure 10: Cancer mortality rate and aerobic physical activity across European countries in 2019.



3.2.2 European Health for All database (HFA-DB)

The European Health for All Database (HFA-DB) is a comprehensive repository of health-related statistics for European countries, maintained by the World Health Organization (WHO) Regional Office for Europe. The database aims to support evidence-based public health policy and decision-making by providing reliable and comparable data on health status, risk factors, and healthcare systems.

HFA-DB includes a wide range of indicators related to morbidity, mortality, health determinants, and healthcare utilisation, facilitating cross-national comparisons. It plays a vital role in monitoring public health trends and addressing health disparities among European populations. By offering information on various health determinants, HFA-DB supports efforts to improve health outcomes and advance health equity across Europe.

The following figures present data on cancer incidence and mortality rates across Europe, specifically focusing on female breast cancer and respiratory tract cancer. Each figure includes the following metrics for various variables: "Incidence of female breast cancer per 100,000 in Europe," "Mortality for breast cancer per 100,000 in Europe," "New cases of respiratory tract cancer in Europe," "Number of new cases of breast cancer in Europe," "Incidence for respiratory tract cancer per 100,000 in Europe," and "Mortality for respiratory tract cancer per 100,000 in Europe." The variables selected include Alcohol consumption, Tobacco smoking, BMI - Overweight, BMI - Obesity, Nutrition - Cereal per year, Nutrition - Calories per day, Nutrition - Fruits and Vegetables per person, Nutrition - Fat available per person, Water supply, Sewage system, Unemployment rate and incidence of mental disorders.

Figure 11 presents six bivariate maps illustrating the association between various cancer metrics and alcohol consumption across European countries. The first map on the left displays the incidence of female breast cancer per 100,000 population. Countries such as France, Belgium, Germany, and Hungary exhibit the darkest shades, indicating high incidence rates of breast cancer and alcohol consumption.

The second map on the left illustrates mortality rates for breast cancer per 100,000 population. Here, Hungary stands out with the darkest colouring, accompanied by Slovenia, Ireland, and Latvia, reflecting elevated mortality rates.

The third map on the left presents the number of new cases of respiratory tract cancer. Hungary, Austria, Poland, Germany, Belgium, and France show the highest values. Following this trend, the first map on the right details the number of new cases of breast cancer in Europe. Again, Hungary, Germany, Poland, Belgium, and France present high case counts.

The second map on the right illustrates the incidence of respiratory tract cancer per 100,000 population. In this case, Estonia, Hungary, Slovenia, Germany, Poland, Belgium, and France present high incidence rates. Finally, the third map on the right shows mortality rates for respiratory tract cancer per 100,000 population. Hungary, Slovenia, Poland, and Belgium emerge as regions with the highest mortality rates.

Figure 11: Cancer metrics and alcohol consumption in Europe.

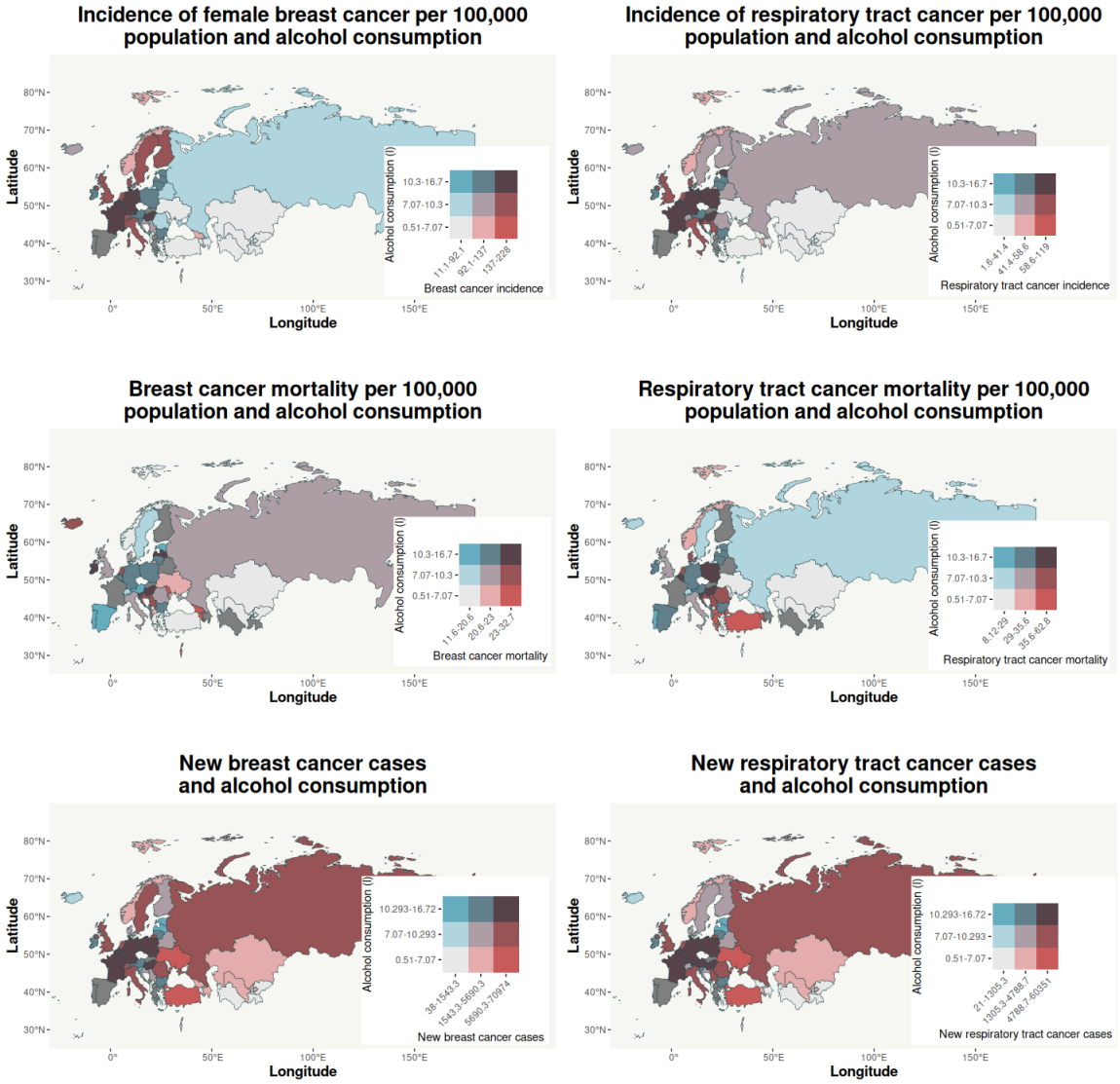


Figure 12 illustrates the connection between cancer metrics and tobacco use across Europe. The first map on the left highlights the incidence of female breast cancer per 100,000 population, with Hungary and France showing the highest rates. The second map on the left depicts mortality rates for breast cancer per 100,000 population, where Hungary, Croatia, Serbia, Latvia, and Georgia exhibit elevated mortality.

The third map on the left presents new cases of respiratory tract cancer, identifying France, Hungary, Serbia, and Turkey as countries with significant case numbers. The first map on the right details the number of new breast cancer cases, again highlighting Hungary, France, and Turkey as regions of concern.

In the second map on the right, the incidence of respiratory tract cancer per 100,000 population is shown, with high rates in France, Hungary, Croatia, Serbia, and Estonia. The final right map illustrates mortality rates for respiratory tract cancer per 100,000 population, noting high mortality in Hungary, Croatia, Serbia, Greece, and Turkey.

Figure 12: Cancer metrics and tobacco use in Europe.

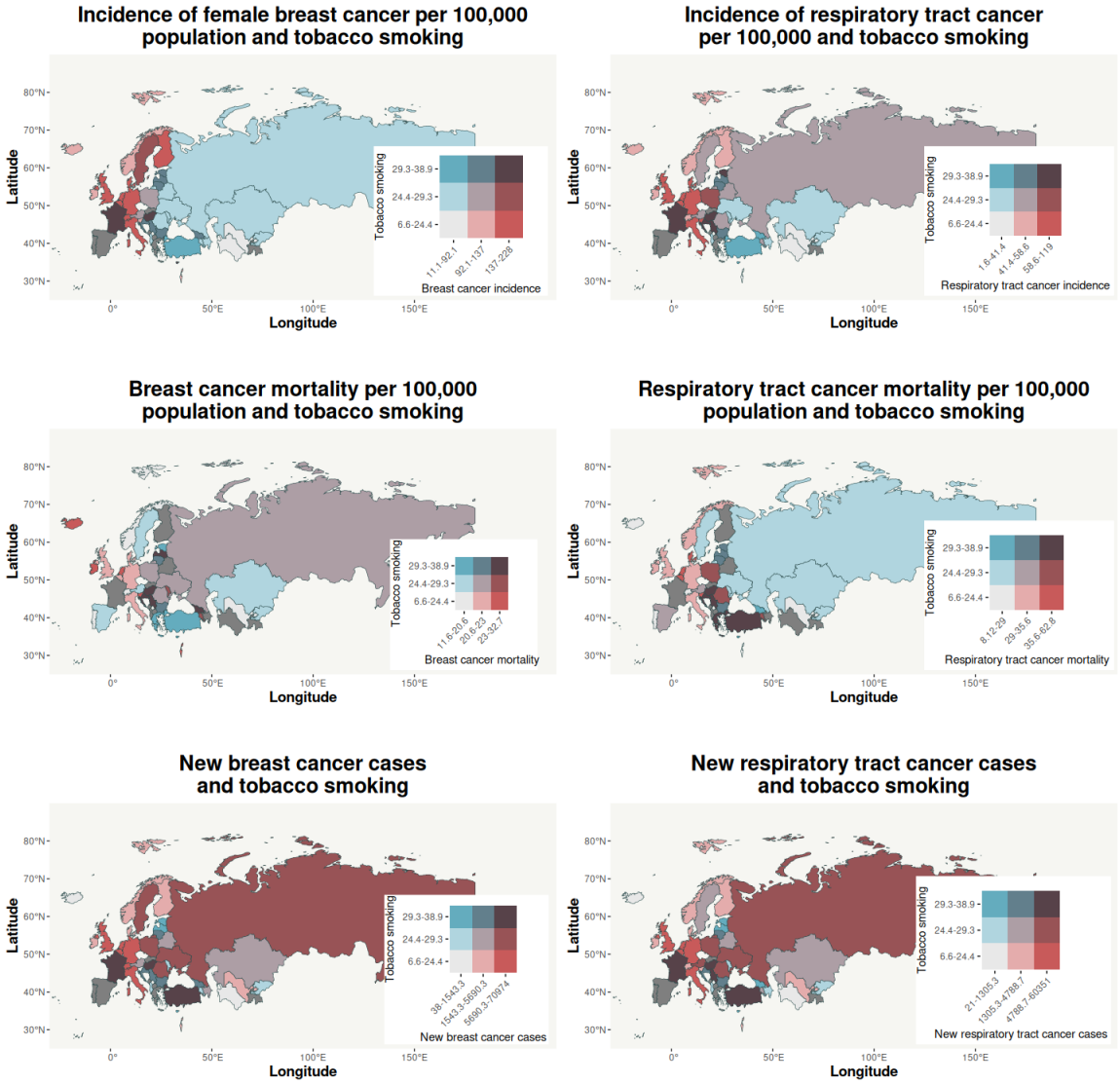


Figure 13 illustrates the connection between cancer metrics and overweight BMI across Europe. The first map on the left highlights the incidence of female breast cancer per 100,000 population, with Hungary, France, and the United Kingdom showing the highest rates. The second map on the left depicts mortality rates for breast cancer per 100,000 population, where Hungary, Croatia, and Ireland exhibit elevated mortality figures.

The third map on the left presents new cases of respiratory tract cancer, identifying Hungary, France, the United Kingdom, and Turkey as countries with significant case numbers. Similarly, the first map on the right details the number of new breast cancer cases, again highlighting Hungary, France, the United Kingdom, and Turkey as regions of concern.

In the second map on the right, the incidence of respiratory tract cancer per 100,000 population is shown, with high rates in Hungary, France, and Turkey. Finally, the third map on the right illustrates mortality rates for respiratory tract cancer per 100,000 population, noting high mortality in Hungary, Croatia, Belgium, Greece, and Turkey.

Figure 13: Cancer metrics and overweight BMI in Europe.

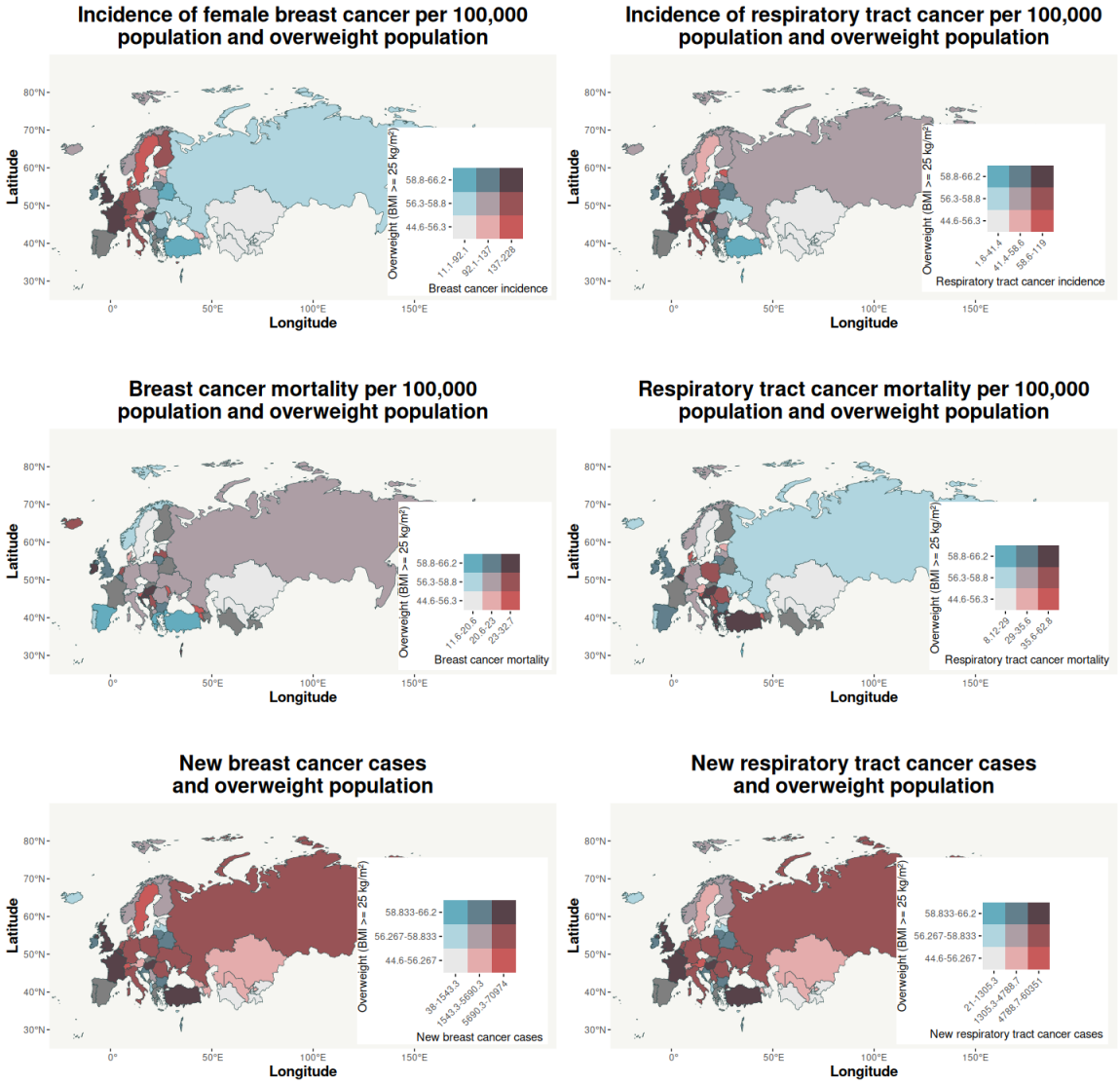


Figure 14 illustrates the connection between cancer metrics and obesity across Europe. The first map on the left highlights the incidence of female breast cancer per 100,000 population, with Hungary and the United Kingdom showing the highest rates. The second map on the left depicts mortality rates for breast cancer per 100,000 population, where Hungary, Croatia, Latvia, and Ireland exhibit elevated mortality figures.

The third map on the left presents new cases of respiratory tract cancer, identifying Hungary, the United Kingdom, Ukraine, and Turkey as countries with significant case numbers. The first map on the right details the number of new breast cancer cases, again highlighting Hungary, the United Kingdom, Ukraine, and Turkey as regions of concern.

In the second map on the right, the incidence of respiratory tract cancer per 100,000 population is shown, with high rates in Hungary, Croatia, and the United Kingdom. The final right map illustrates mortality rates for respiratory tract cancer per 100,000 population, noting high mortality in Hungary, Croatia, Greece, and Turkey.

Figure 14: Cancer metrics and obesity in Europe.

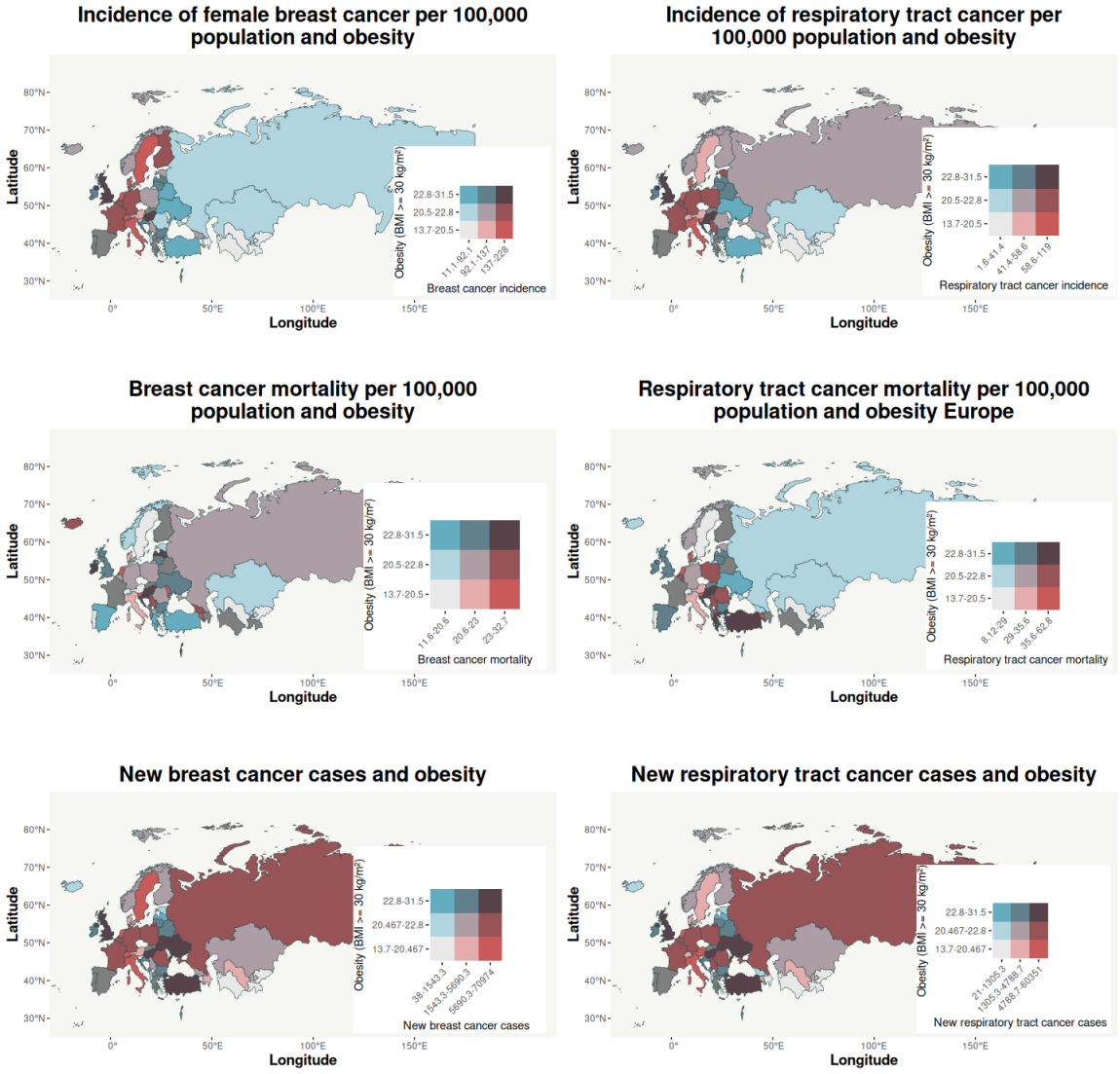


Figure 15 depicts the relationship between cancer metrics and cereal consumption per year in Europe. The first map on the left highlights the incidence of female breast cancer per 100,000 population, with Italy and France reporting the highest rates. The second map on the left illustrates mortality rates for breast cancer per 100,000 population, where Moldova and Georgia show significantly elevated mortality figures.

The third map on the left presents new cases of respiratory tract cancer, identifying Italy, France, Russia, Ukraine, Romania, and Turkey as countries with notable case numbers. The first map on the right details the total number of new breast cancer cases, again emphasising Italy, France, Russia, Ukraine, Romania, and Turkey as areas of concern.

In the second map on the right, the incidence of respiratory tract cancer per 100,000 population is displayed, with high rates observed in Italy and France. Finally, the third map on the right shows mortality rates for respiratory tract cancer per 100,000 population, indicating higher mortality in Romania and Turkey.

Figure 15: Cancer metrics and cereal consumption in Europe.

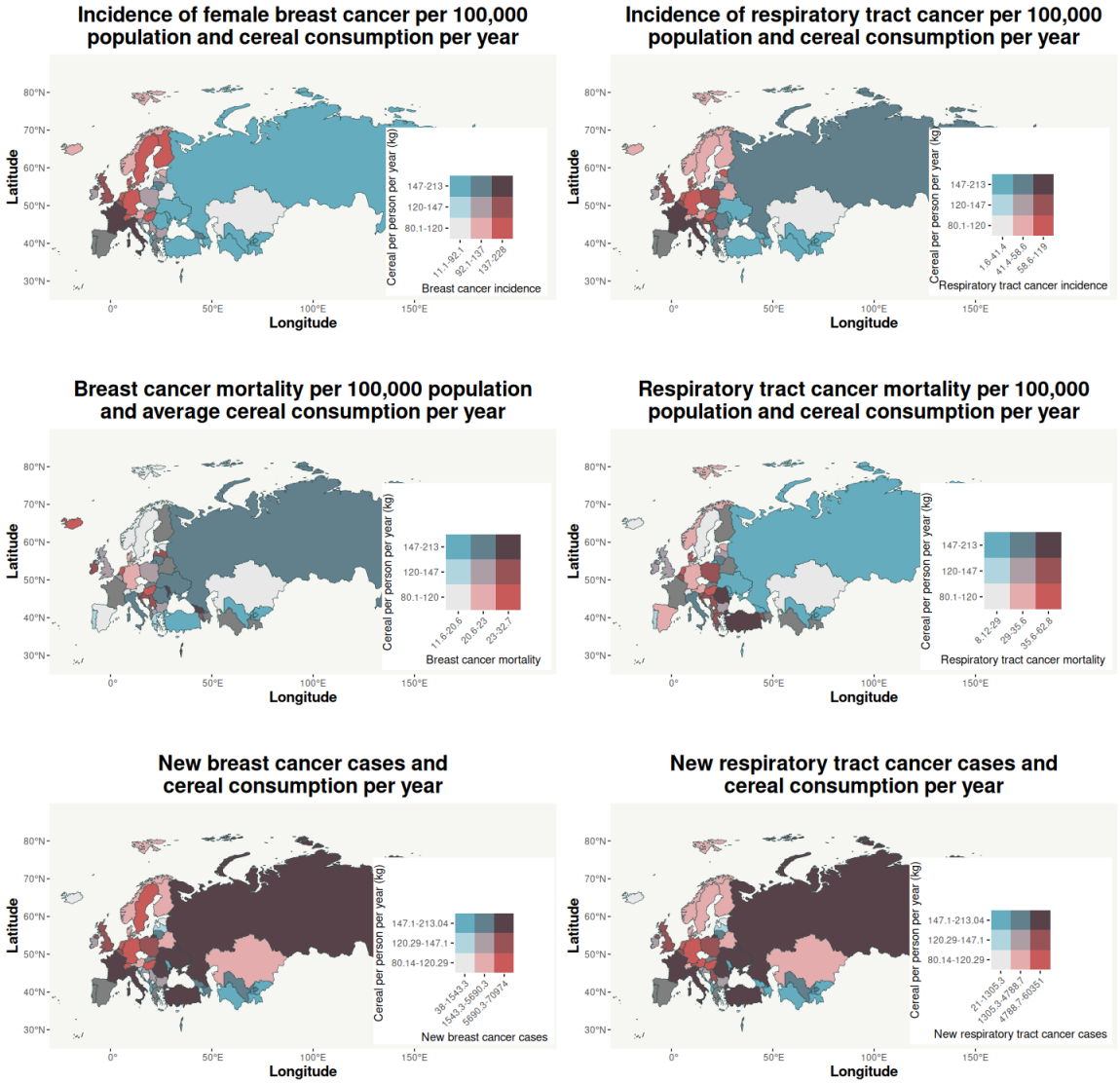


Figure 16 illustrates the association between cancer metrics and average daily calorie intake across Europe. The first map on the left showcases the incidence of female breast cancer per 100,000 population, with Italy, France, Luxembourg, Germany, Belgium, and Switzerland exhibiting the highest rates. The second map on the left represents mortality rates for breast cancer per 100,000 population, where Iceland and Ireland reveal notably high mortality figures.

The third map on the left presents new cases of respiratory tract cancer, highlighting Italy, France, Belgium, Austria, Romania, and Turkey as countries with significant case counts. The first map on the right details the total number of new breast cancer cases, again pointing to Italy, France, Germany, Belgium, Switzerland, Romania, and Turkey as areas of concern.

In the second map on the right, the incidence of respiratory tract cancer per 100,000 population is shown, with elevated rates in Italy, France, Germany, and Belgium. The final map on the right illustrates mortality rates for respiratory tract cancer per 100,000 population, noting high mortality in Belgium, Romania, Greece, and Turkey.

Figure 16: Cancer metrics and calories per day in Europe.

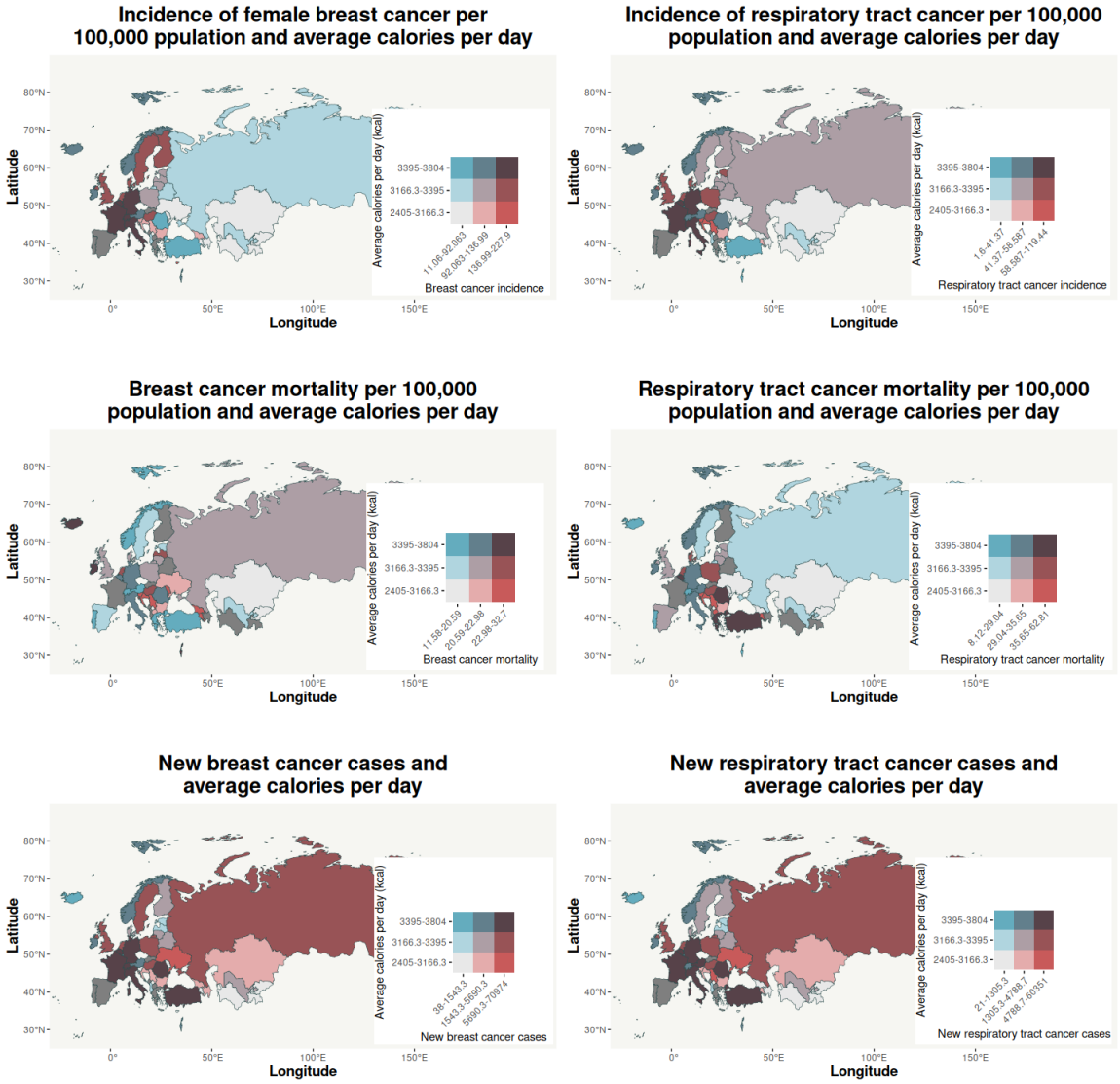


Figure 17 depicts the relationship between cancer metrics and the consumption of fruits and vegetables per person across Europe. The first map on the left highlights the incidence of female breast cancer per 100,000 population, with Italy and Belgium reporting the highest rates. The second map on the left illustrates mortality rates for breast cancer per 100,000 population, where Armenia shows notably high mortality.

The third map on the left presents new cases of respiratory tract cancer, identifying Italy, Belgium, Romania, and Turkey as countries with significant case numbers. The first map on the right details the total number of new breast cancer cases, again emphasising Italy, Belgium, Romania, and Turkey as areas of concern.

In the second map on the right, the incidence of respiratory tract cancer per 100,000 population is displayed, with high rates observed in Italy and Belgium. The final map on the right shows mortality rates for respiratory tract cancer per 100,000 population, indicating elevated mortality in Belgium, Romania, Greece, and Turkey.

Figure 17: Cancer metrics and fruit and vegetable consumption per person in Europe.

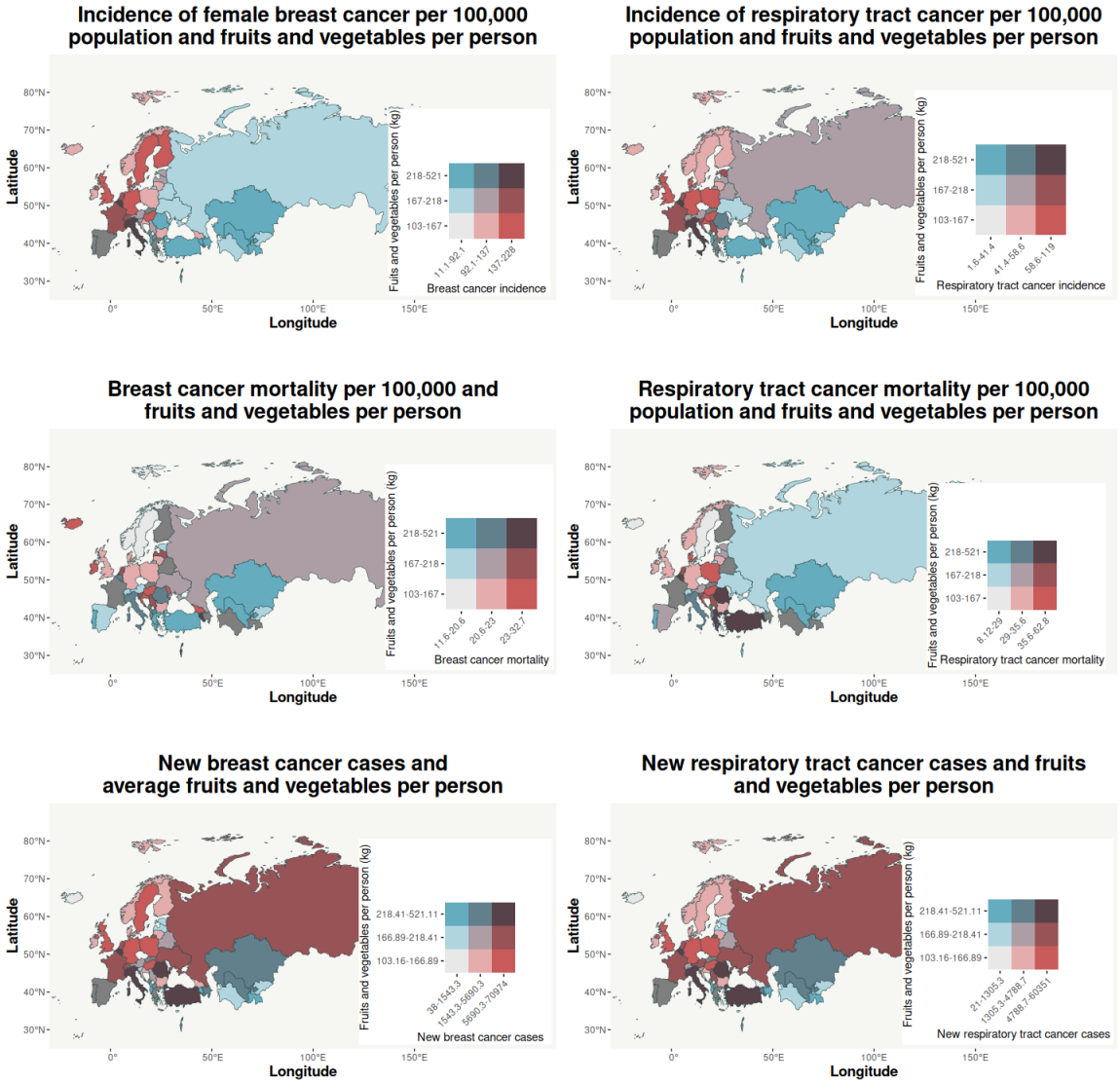


Figure 18 illustrates the association between cancer metrics and fat availability per person across Europe. The first map on the left highlights the incidence of female breast cancer per 100,000 population, with Finland, France, Italy, Belgium, the United Kingdom, Luxembourg, Switzerland, and Hungary reporting the highest rates. The second map on the left shows mortality rates for breast cancer per 100,000 population, where Iceland, Ireland, and Hungary exhibit notably high mortality figures.

The third map on the left presents new cases of respiratory tract cancer, identifying France, Italy, Belgium, Germany, Austria, and Hungary as countries with significant case counts. Similarly, the first map on the right details the total number of new breast cancer cases, emphasising France, Italy, Belgium, the United Kingdom, and Hungary as areas of concern.

In the second map on the right, the incidence of respiratory tract cancer per 100,000 population is displayed, with high rates observed in France, Italy, Belgium, the United Kingdom, and Hungary. The final map on the right illustrates mortality rates for respiratory tract cancer per 100,000 population, indicating elevated mortality in Belgium, Greece, and Hungary.

Figure 18: Cancer metrics and fat available per person in Europe.

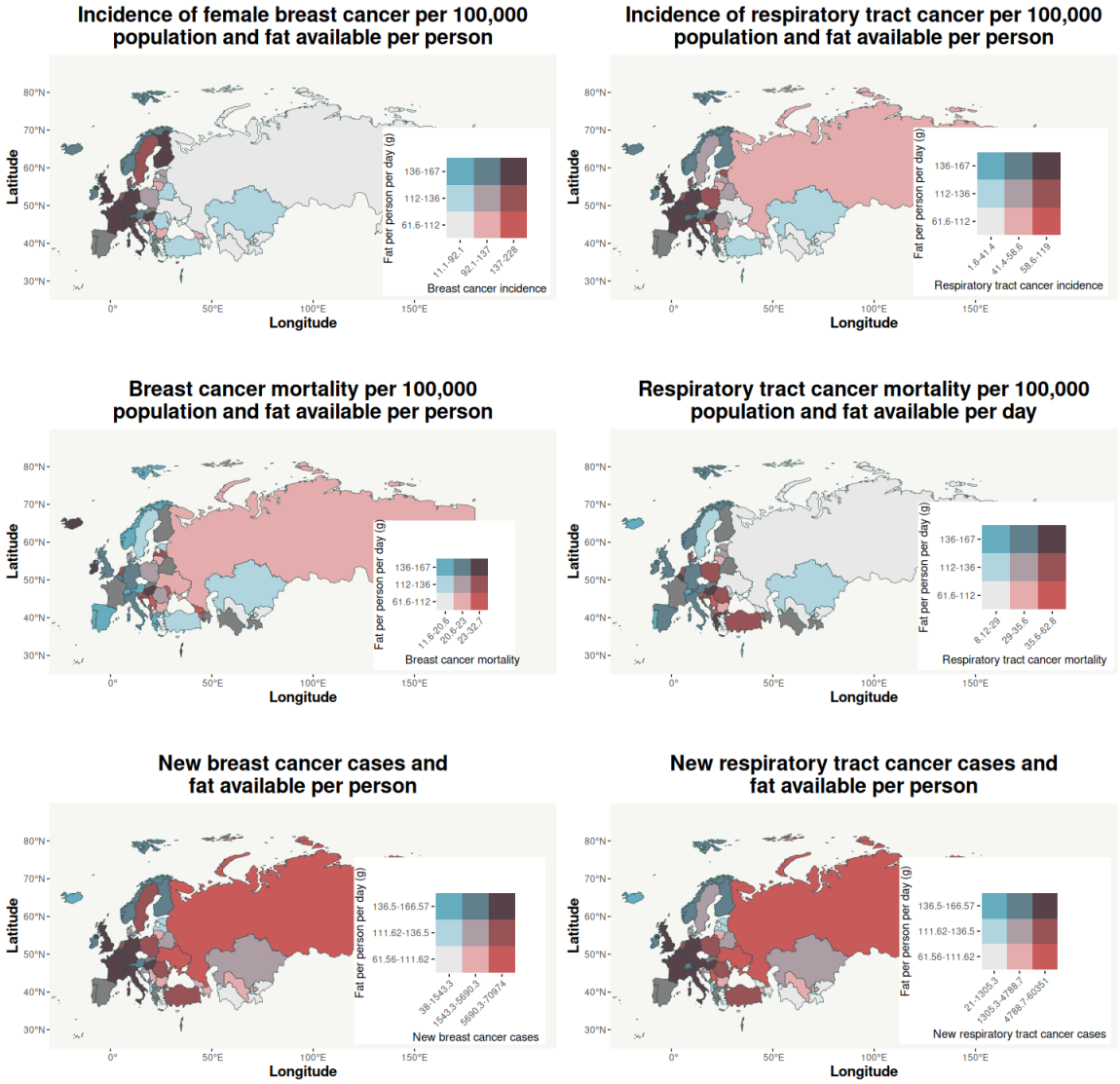


Figure 19 examines the relationship between cancer metrics and access to water supply systems across Europe. The first map on the left highlights the incidence of female breast cancer per 100,000 population, with Finland, Sweden, France, Italy, Belgium, the United Kingdom, and Hungary reporting the highest rates. The second map on the left depicts mortality rates for breast cancer per 100,000 population, where Iceland, Croatia, Serbia, and Armenia show significantly elevated mortality figures.

The third map on the left presents new cases of respiratory tract cancer, identifying France, Germany, the United Kingdom, and Russia as countries with considerable case numbers. The first map on the right details the total number of new breast cancer cases, again emphasising France, Germany, the United Kingdom, and Russia as areas of concern.

In the second map on the right, the incidence of respiratory tract cancer per 100,000 population is shown, with notable rates in Hungary and Denmark. Finally, the third map on the right illustrates mortality rates for respiratory tract cancer per 100,000 population, indicating higher mortality in Serbia and Poland.

Figure 19: Cancer metrics and access to water supply systems in Europe.

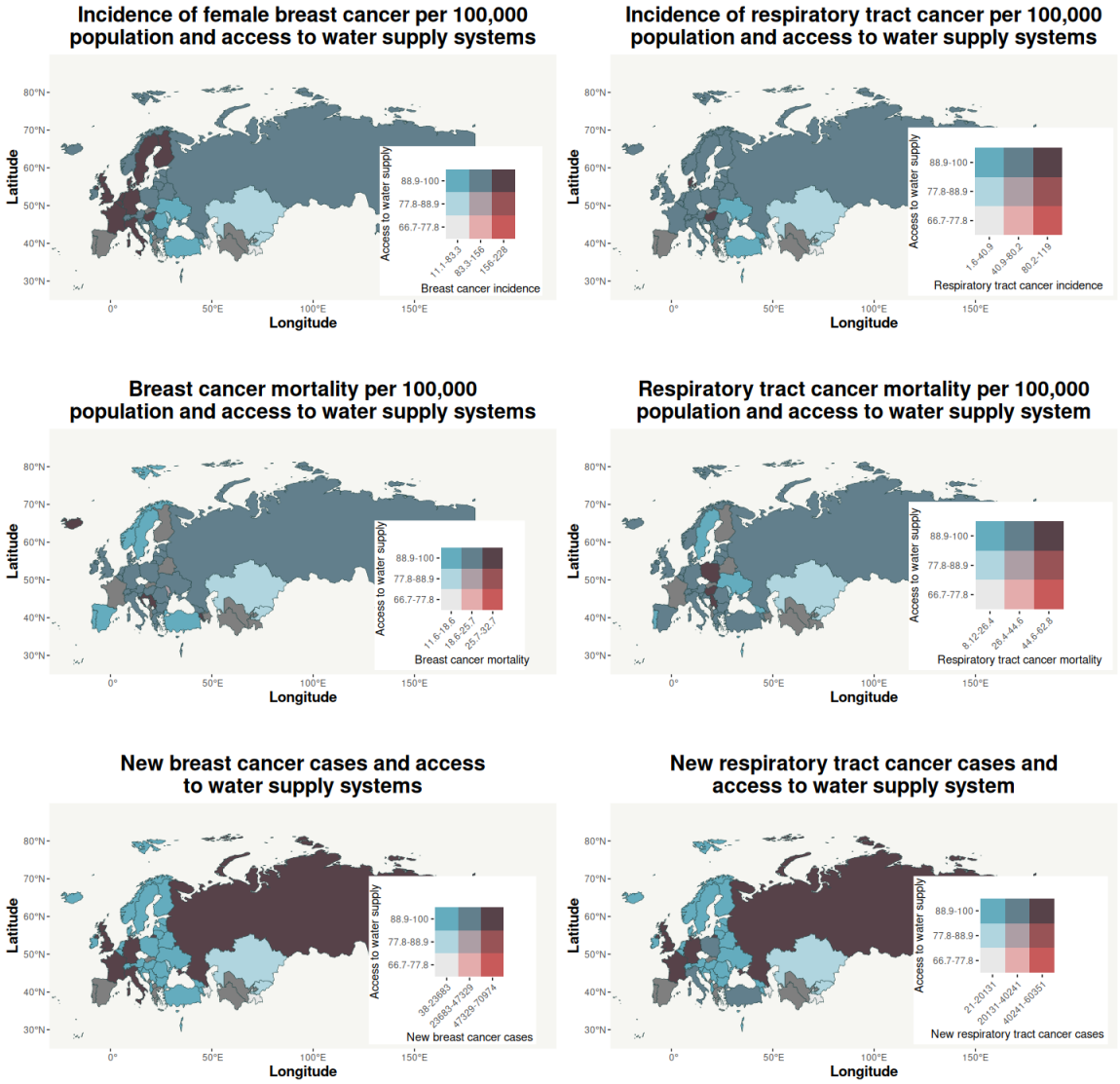


Figure 20 explores the connection between cancer metrics and access to sewage systems across Europe. The first map on the left highlights the incidence of female breast cancer per 100,000 population, with Finland, Sweden, Denmark, Germany, Italy, Belgium, and Switzerland reporting the highest rates. The second map on the left depicts mortality rates for breast cancer per 100,000 population. In contrast, the third map on the left presents new cases of respiratory tract cancer, identifying France, Germany, the United Kingdom, and Russia as countries with significant case numbers.

The first map on the right details the total number of new breast cancer cases, emphasising Italy, Germany, Belgium, Hungary, Switzerland, and Sweden as areas of concern. In the second map on the right, the incidence of respiratory tract cancer per 100,000 population is shown, with notable rates observed in Italy, Germany, Belgium, and Hungary. The final map on the right illustrates mortality rates for respiratory tract cancer per 100,000 population, indicating higher mortality in Denmark and Belgium.

Figure 20: Cancer metrics and access to the sewage system in Europe.

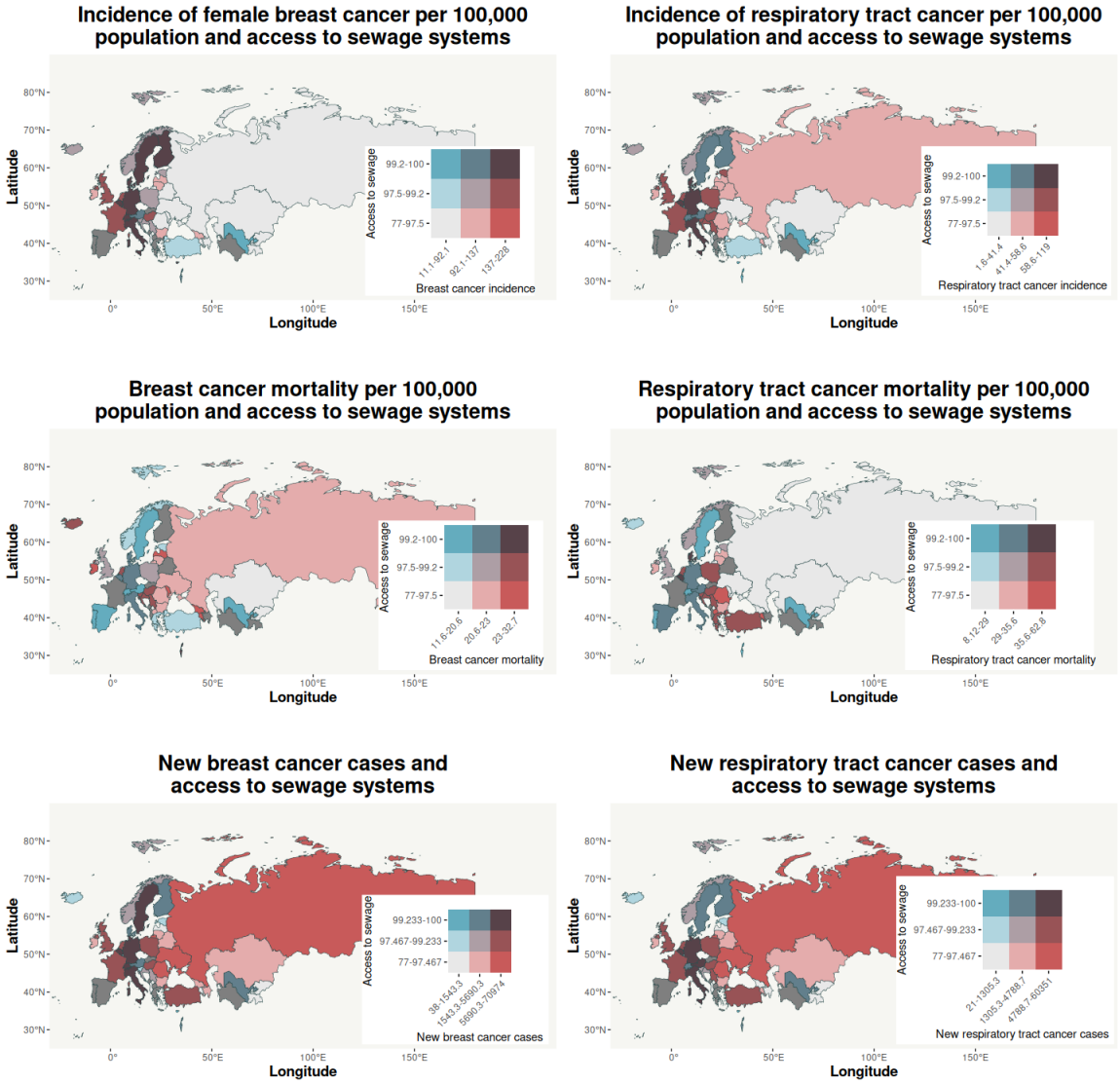


Figure 21 examines the relationship between cancer metrics and the unemployment rate across Europe. The first map on the left highlights the incidence of female breast cancer per 100,000 population, with Finland, Sweden, Denmark, Germany, Italy, and Belgium showing the highest rates. The second map on the left depicts mortality rates for breast cancer per 100,000 population. In contrast, the third map on the left presents new cases of respiratory tract cancer, identifying Germany, Italy, Belgium, and Austria as countries with significant case numbers.

The first map on the right details the total number of new breast cancer cases, again emphasising Italy, Germany, Belgium, Switzerland, and Sweden as areas of concern. In the second map on the right, the incidence of respiratory tract cancer per 100,000 population is displayed, with notable rates in Denmark, Italy, Germany, and Belgium. The final map on the right illustrates mortality rates for respiratory tract cancer per 100,000 population, indicating higher mortality figures in Denmark and Belgium.

Figure 21: Cancer metrics and unemployment rate in Europe.

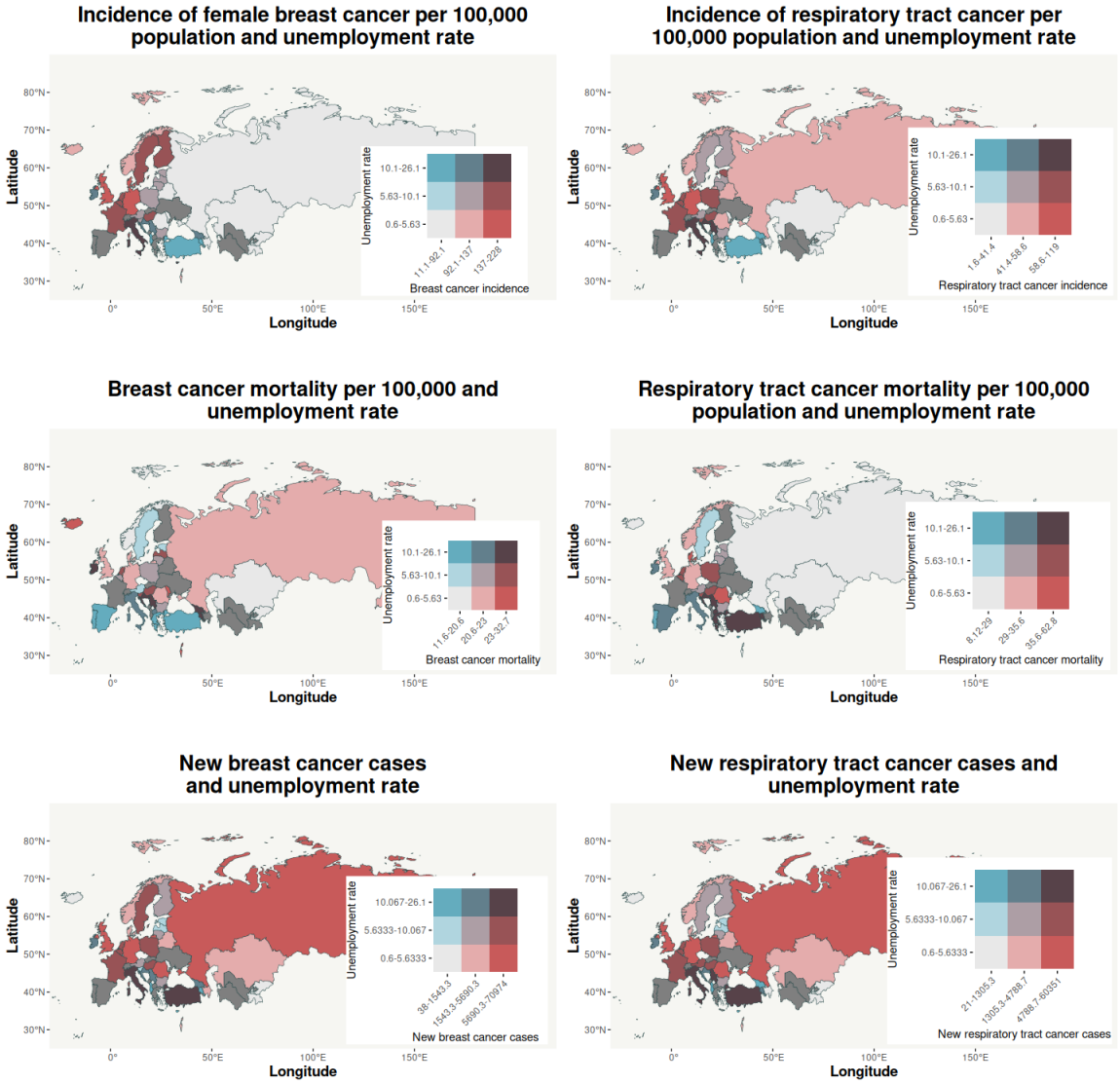
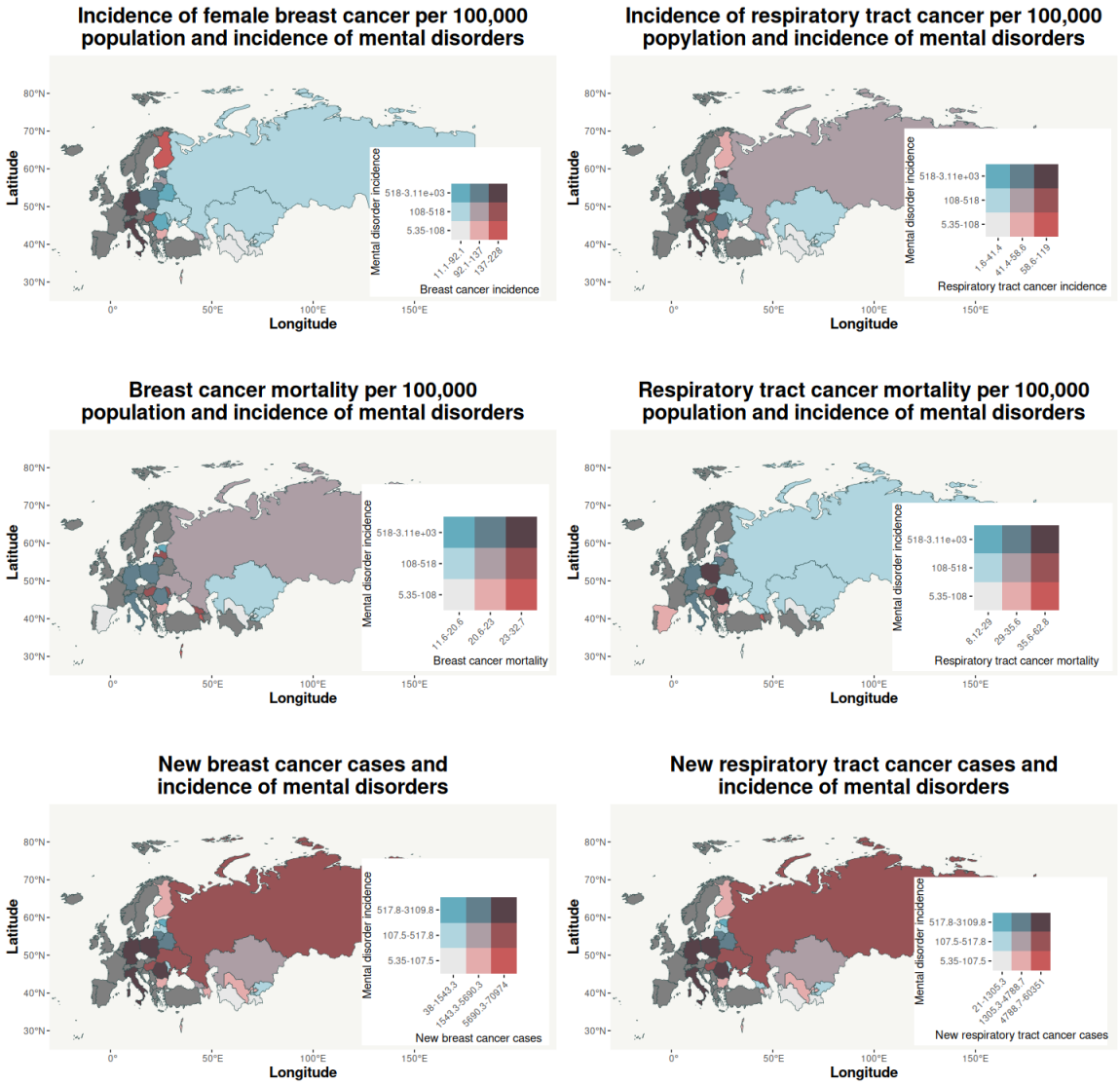


Figure 22 explores the association between cancer metrics and the incidence of mental disorders across Europe. The first map on the left highlights the incidence of female breast cancer per 100,000 population, with Germany and Italy reporting the highest rates. The second map on the left depicts mortality rates for breast cancer per 100,000 population. In contrast, the third map on the left presents new cases of respiratory tract cancer, identifying Germany, Poland, Romania, and Italy as countries with significant case numbers.

The first map on the right details the total number of new breast cancer cases, again emphasising Germany, Poland, Romania, and Italy. In the second map on the right, the incidence of respiratory tract cancer per 100,000 population is shown, with notable rates observed in Germany, Poland, Italy, and Estonia. The final map on the right illustrates mortality rates for respiratory tract cancer per 100,000 population, indicating higher mortality figures in Poland and Romania.

Figure 22: Cancer metrics and incidence of mental disorders in Europe.



4. Conclusions

This report synthesises the extensive data collection and analysis conducted under Task 3.2 of the iBeChange project. This research focused on identifying and analysing publicly available data related to psychosocial and lifestyle risk factors for cancer, spanning various European countries. The process involved a careful data mining approach, which can help understand the connections between psychosocial determinants, lifestyle choices and cancer outcomes across diverse populations.

Key findings of the analyses indicate that lifestyle factors such as diet, smoking, and exercise can have profound implications on cancer risks and general health outcomes. Specifically, it was observed that:

- Smoking and high alcohol consumption are significantly associated with increased cancer risks.
- Dietary habits, particularly low fruit and vegetable intake, correlate with higher incidences of certain cancers.
- Regular physical activity appears to mitigate some of the risk factors associated with sedentary lifestyle diseases, including cancer.

Moreover, the analysis underscores significant psychosocial factors such as socio-economic status, education level, employment status, social support, and mental health conditions. These elements influence health behaviours and subsequently impact cancer risk and survivorship. Lower socio-economic and educational levels are associated with higher cancer risks and poorer health outcomes, highlighting the need for interventions that consider the broader socio-psychological environment. Employment status and mental health conditions like depression and anxiety, prevalent among individuals with chronic illnesses, including cancer, also significantly affect health behaviours and outcomes.

Throughout this study, data from 28 diverse databases were successfully documented and standardised. For detailed analysis, five databases were strategically selected: three containing individual-level data and two with aggregate data. This selection was driven by the objective to thoroughly explore trends and disparities in cancer prevalence, alongside associated psychosocial and behavioral factors, encompassing both detailed individual observations and broader geographic patterns. The integration of these data using the OMOP Common Data Model facilitated a robust analysis, revealing trends and disparities in cancer prevalence and associated behavioural factors.

These efforts contribute to the iBeChange project’s overarching goal of developing a personalised prevention tool. The insights from this analysis will inform the development of targeted interventions designed to mitigate these risks by promoting healthier lifestyle choices among European populations. Specifically, the insights gained from this task will directly influence the tasks within Work Package 4 (WP4), which focuses on the development of the reinforcement learning-based personalised behavioural change platform. Part of the publicly available data gathered in Task 3.2, specifically the individual-level data, will be used to train the reinforcement learning models in Task 4.5. Additionally, the information derived here will contribute to the tailoring of interventions to address the specific needs and risk profiles, ensuring that the iBeChange intervention is based on solid evidence. Furthermore, the findings will support tasks in Work Package 5 (WP5), aimed at evaluating the effectiveness of these interventions in real-world settings by providing baseline data and metrics for comparison. This ensures a coherent flow of information across the project, enhancing the overall impact and relevance of the iBeChange platform in advancing primary cancer prevention.

5. References

1. Kovalchik S. RISmed: Download Content from NCBI Databases, https://cran.r-project.org/web/packages/RISmed/index.html (2021, accessed 1 November 2024).
2. Informatics OHDS and. *The Book of OHDSI*, https://ohdsi.github.io/TheBookOfOhdsi/ (accessed 1 November 2024).
3. OMOP Common Data Model – OHDSI, https://www.ohdsi.org/data-standardization/the-common-data-model/ (accessed 17 September 2021).
4. Athena, https://athena.ohdsi.org/search-terms/start (accessed 1 November 2024).
5. Kuhn M, Johnson K. Applied Predictive Modeling. 2013.
6. Faraone SV. Interpreting estimates of treatment effects: implications for managed care. *P T Peer-Rev J Formul Manag* 2008; 33: 700–11.
7. Ipsos | Global Market Research and Public Opinion Specialist, https://www.ipsos.com/en-uk (2024, accessed 29 October 2024).
8. Healthy Ireland Survey, https://www.ucd.ie/issda/data/healthyireland/healthyirelandsurvey2022/ (2022).
9. Matthews S, Pierce M, O’Brien Green S, et al. *Dying and death in Ireland: what do we routinely measure, how can we improve.* Report, Dublin: Irish Hospice Foundation, https://www.drugsandalcohol.ie/35234/ (November 2021, accessed 29 October 2024).
10. Archive ISSD. Home, Irish Social Science Data Archive. *Irish Social Science Data Archive*, https://www.ucd.ie/issda/data/healthyireland/healthyirelandsurvey2022/ (accessed 29 October 2024).
11. Central Statistics Office. *data.cso.ie*. Ireland, https://data.cso.ie/ (accessed 11 January 2024).
12. Background Notes Irish Health Survey 2019 - Main Results - Central Statistics Office, https://www.cso.ie/en/releasesandpublications/ep/p-ihsmr/irishhealthsurvey2019-mainresults/backgroundnotes/ (accessed 29 October 2024).
13. Irish Health Survey (IHS) - CSO - Central Statistics Office, https://www.cso.ie/en/methods/health/irishhealthsurveyihs/ (accessed 29 October 2024).
14. EUROSTAT. EU statistics on income and living conditions, https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions (accessed 11 January 2024).
15. EPIC: About EPIC, https://epic.iarc.fr/about/ (accessed 30 October 2024).
16. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002; 5: 1113–1124.
17. EUROSTAT. European health interview survey, https://ec.europa.eu/eurostat/web/microdata/european-health-interview-survey.
18. EUROSTAT. Causes of death statistics, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Causes_of_death_statistics.
19. Vucenik I, Stains JP. Obesity and cancer risk: evidence, mechanisms, and recommendations. *Ann N Y Acad Sci* 2012; 1271: 37–43.
20. Gandini S, Botteri E, Iodice S, et al. Tobacco smoking and cancer: a meta-analysis. *Int J Cancer* 2008; 122: 155–164.

6. Appendices

6.1 Appendix 1: Source code for the pubmed Database Search R function.

```
#' Database Search Report
#'
#' @description This function searches PubMed for databases based on
a given query and extracts relevant information,
#' including titles, authors, affiliations, dates, and snippets from
abstracts.
#'
#' @param query Search query.
#' @param base_terms Set of key terms used to find specific databases
(default: registry, database, dataset, claim, cohort, and trial).
#' @param include_base_terms Logical flag to decide whether to include
base_terms in the search (default: TRUE).
#' @param remove_terms Set of terms to be removed from the search
(default: NULL).
#' @param countries Set of countries to be included in the search
(default: NULL).
#' @param nmax Maximum number of records to retrieve (default: 1000).
#' @param nwords The number of words surrounding the key terms in the
snippets (default: 10).
#' @param filename The name of the file to save the results. If NULL,
a filename is automatically generated (default: NULL).
#' @param silent Logical flag to suppress messages during the function
execution (default: FALSE).
#'
#' @export
pubmedDatabaseSearch <- function(query, base_terms = c("registry",
"database", "dataset", "claim", "cohort", "trial"),
                                 remove_terms = NULL, countries =
NULL,
```

```
                                    include_base_terms = TRUE, nmax =
1000, nwords = 10, filename = NULL, silent = FALSE) {
  # If no filename was provided, generate one randomly based on the
inputs using SHA256 hash
  if (is.null(filename)) {
    filename <- sprintf("database_search_%s.csv", digest(c(query,
base_terms, nmax), algo = "sha256"))
  } else {
    ext <- tools::file_ext(filename)
    if (ext == "") {
      filename <- sprintf("%s.csv", filename)
    } else {
      assert_subset(ext, c("csv"), empty.ok = FALSE)
    }
  }

  if (include_base_terms) query <- sprintf("(%s) AND (%s)", query,
paste0(base_terms, collapse = " OR "))

  if (!is.null(countries)) query <- sprintf("(%s) AND (%s)", query,
paste0(paste0(countries, "[AD]"), collapse = " OR "))

  silence <- if (silent) function(x) suppressMessages(x) else
function(x) x

  silence(message(sprintf("Final query: %s", query)))

  silence(message("Retrieving PubMed records... This may take a
while"))
  pubmed_summary <- RISmed::EUtilsSummary(query, retmax = nmax, db =
"pubmed")
```

```
  pubmed_records <- RISmed::EUtilsGet(pubmed_summary)


  silence(message("Processing records..."))


  # Extracting and processing authors information
  authors <- lapply(RISmed::Author(pubmed_records), function(author_list) {
    if (invalid(author_list)) {
      data.frame(author_id = numeric(0), ForeName = character(0), LastName = character(0))
    } else {
      df <- data.frame(author_id = author_list$order, ForeName = author_list$ForeName, LastName = author_list$LastName)
      df[!duplicated(df[, c("ForeName", "LastName")]), ]
    }
  })


  # Extracting and processing affiliations information
  affiliations <- lapply(RISmed::Affiliation(pubmed_records), function(aff_list) {
    if (invalid(aff_list)) {
      data.frame(affiliation = character(0), author_id = numeric(0), affili_id = numeric(0))
    } else {
      data.frame(
        affiliation = as.character(aff_list),
        author_id = as.numeric(names(aff_list)),
        affili_id = as.numeric(factor(aff_list, levels = unique(aff_list)))
      )
```

```
    }
  })

  # Merging authors and affiliations data
  authors_affiliations <- mapply(function(authors_df, affilis_df) {
    merge(authors_df, affilis_df, by = "author_id", all.x = TRUE)
  }, authors, affiliations, SIMPLIFY = FALSE)

  # Constructing authors' names string
  authors_names <- sapply(authors_affiliations, function(auths_affs)
{
    if (nrow(auths_affs) == 0) {
      ""
    } else {
      df <- aggregate(affili_id ~ ForeName + LastName + author_id,
auths_affs, FUN = list, na.action = identity)
      affsids <- sapply(df$affili_id, function(x) {
        if (all(is.na(x))) ""
        else paste0(" [", paste(x, collapse = ","), "]")
      })
      paste(paste0(df$ForeName, " ", df$LastName, affsids), collapse
= "; ")
    }
  })

  # Constructing affiliations names string
  affiliations_names <- sapply(affiliations, function(affs) {
    affs <- affs[!duplicated(affs$affili_id), ]
    if (nrow(affs) > 0) {
```

```
      paste(paste0("[", affs$affili_id, "] ", affs$affiliation),
collapse = "; ")
    } else {
      ""
    }
  })

  # Coalesce diferent sources of dates
  dates <- dplyr::coalesce(
    as.Date(paste(RISmed::YearPubDate(pubmed_records),
                  RISmed::MonthPubDate(pubmed_records),
                  RISmed::DayPubDate(pubmed_records), sep = "-"),
format = "%Y-%b-%d"),
    as.Date(paste(RISmed::YearArticleDate(pubmed_records),
                  RISmed::MonthArticleDate(pubmed_records),
                  RISmed::DayArticleDate(pubmed_records), sep = "-"),
format = "%Y-%m-%d"),
    as.Date(paste(RISmed::YearPubmed(pubmed_records),
                  RISmed::MonthPubmed(pubmed_records),
                  RISmed::DayPubmed(pubmed_records), sep = "-"),
format = "%Y-%m-%d")
  )

  # Creating links to the PubMed articles
  links <- paste0("https://pubmed.ncbi.nlm.nih.gov/",
RISmed::PMID(pubmed_records), "/")

  # Processing abstract texts and removing unwanted tags
  abstracts <- RISmed::AbstractText(pubmed_records)
```

```
  abstracts <- stringr::str_replace_all(abstracts, "((Label|NlmCategory)=\"[^\\\"]+\"\\s*)+:?", "")
  abstracts <- stringr::str_replace_all(abstracts, "(^\\s+|\\s+$|(?<=\\s)\\s+)", "")

  # Extracting snippets from abstracts based on the specified terms: nwords before + term(suffix)? + nwords after
  nwords <- max(1, nwords)
  snippet_pattern <- paste0("(?:\\w+[\\W]+){0,", nwords, "}(?:", paste(base_terms, collapse = "|"), ")(?:[^\\s]+)?(?:[\\W]+\\w+){0,", nwords, "}")

  snippets <- stringr::str_extract_all(abstracts, snippet_pattern)
  snippets <- sapply(snippets, function(s) {
    if (length(s) > 0) {
      paste(paste0("(", seq_along(s), ") ", s), collapse = "\n")
    } else {
      NA
    }
  })

  # Extracting and processing titles information
  article_titles <- RISmed::ArticleTitle(pubmed_records)

  # Remove records based on the specified terms
  if (!is.null(remove_terms)) {

    matched_titles <- stringr::str_detect(article_titles, stringr::regex(paste0(remove_terms, collapse = "|"), ignore_case = TRUE))
```

```
    matched_abstracts <- stringr::str_detect(abstracts,
stringr::regex(paste0(remove_terms, collapse = "|"), ignore_case =
TRUE))
    matched_entries <- matched_titles | matched_abstracts

    article_titles <- article_titles[!matched_entries]
    snippets <- snippets[!matched_entries]
    authors_names <- as.character(authors_names)[!matched_entries]
    affiliations_names <-
as.character(affiliations_names)[!matched_entries]
    dates <- dates[!matched_entries]
    links <- links[!matched_entries]
  }

  results <- data.frame(
    Title = article_titles,
    Snippets = snippets,
    Authors = authors_names,
    Affiliations = affiliations_names,
    Date = dates,
    Link = links
  )

  silence(message(sprintf("Saving %s...", filename)))
  write.csv(results, file = filename, row.names = FALSE, fileEncoding
= "UTF-8")

  silence(message("Done."))

}
```

6.2 Appendix 2: List of databases selected.

Database	Link	Example of variables available	Data Level	Location	Requirements
European Prospective Investigation into Cancer and Nutrition (EPIC)	https://epic.iarc.fr/	Dietary habits, lifestyle factors, anthropometric measurements, biochemical markers, and cancer incidence.	Individual	10 European countries	Submission of a formal application detailing the study's objectives and methodology. Approval from the EPIC Steering Committee is required.
European Health Interview Survey (EHIS)	https://ec.europa.eu/eurostat/cache/metadata/en/hlth_det_esms.htm	Health status, healthcare use, health determinants, and socio-demographic information.	Individual and regional	European Union	Access is granted to researchers and institutions upon request, following Eurostat's microdata access procedures. Regional-level data is available without restriction.
Global Burden of Disease Study 2019 (GBD 2019)	https://ghdx.healthdata.org/record/ihme-data/gbd-2019-risk-factors-attributable-cancer-burden-estimates	Mortality, morbidity, risk factors, and health metrics across diseases and injuries.	Country and regional	Worldwide	Data is available freely for non-commercial use.
OECD (Organisation for Economic Co-operation and Development)	https://stats.oecd.org/	Economic indicators, health statistics, education data, and social indicators.	Country	OECD Countries	Data is publicly accessible via the OECD.Stat portal.
Breast Cancer Association Consortium (BCAC)	https://bcac.ccge.medschl.cam.ac.uk/	Genetic data, epidemiological information, and clinical outcomes related to breast cancer.	Individual	Worldwide	Researchers must submit a data access request outlining the study's purpose and methodology. Approval from the BCAC Data Access Coordination Committee is required.
Swedish National Database of Research and Development	https://snd.gu.se/en/catalogue/search/cancer%20risk%20factors	Cancer risk factors, information on health, social sciences, and humanities.	Individual and regional	Sweden	Access depends on the specific dataset and may require an application and approval process.

Health Survey Microdata sets	https://www.statistik.at/statistiken/bevoelkerung-und-soziales/gesundheit/gesundheitszustand/gesundheitszustand-selbstberichtet	Self-reported health status, health behaviours, and healthcare utilisation.	Individual	Austria	Requires application for access through Statistics Austria, adhering to their data protection guidelines.
Payroll Tax Statistics	https://www.statistik.at/statistiken/volkswirtschaft-und-oeffentliche-finanzen/oeffentliche-finanzen/steuerstatistiken/lohnsteuerstatistik	Income data, tax information, and employment statistics.	Individual	Austria	Access is subject to application and approval by Statistics Austria, ensuring compliance with data protection laws.
Health Survey 2019 (Austria)	https://www.statistik.at/services/tools/services/amdc-mikrodaten-fuer-die-wissenschaft/scientific-use-files#c15273	Health status, lifestyle factors, and healthcare utilisation.	Individual	Austria	An application is required to access Scientific Use Files through Statistics Austria.
The Estonian Health Interview Survey (Estonia)	https://statistika.tai.ee/pxweb/et/Andmebaas/Andmebaas_05Uuringud	Health behaviours, chronic conditions, and healthcare access.	Individual	Estonia	Requires an application to the National Institute for Health Development in Estonia.
Healthy Ireland Survey (Ireland)	https://data.cso.ie/	Health behaviours, mental health, and lifestyle factors.	Individual	Ireland	Data is available through the Irish Social Science Data Archive (ISSDA) upon application.
Monitoring system - Study of habits affecting the health of the	https://www.spkc.gov.lv/lv/veselibu-ietekmejoso-	Health behaviours, lifestyle factors, and preventive measures.	Country and individual	Latvia	Requires an application to the Centre for Disease Prevention and Control of Latvia.

population of Latvia (Latvia)	paradumu-petijumi https://www.spkc.gov.lv/lv/datu-bazes				
Cancer Prevention Statistics (Lithuania)	https://osp.stat.gov.lt/EN/statistiniu-rodikliu-analize?hash=4300a21c-a5c9-4fa7-bc51-7246fb1e51c4#/	Screening participation rates, early detection statistics, and preventive measures.	Country and regional	Lithuania	Publicly accessible through the Lithuanian Department of Statistics portal.
National Health Interview Survey (INS) (Portugal)	https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0008934&xlang=pt&contexto=bd&selTab=tab2	Health status, healthcare utilisation, lifestyle factors, and chronic conditions.	Individual	Portugal	Requires a formal request to the Portuguese National Institute of Statistics (INE).
NIJZ data portal (Slovenia)	https://podatki.nijz.si/pxweb/sl/NIJZ%20podatkovni%20portal/	Health statistics, including disease prevalence, risk factors, and healthcare services.	Country and regional	Slovenia	Publicly accessible through the National Institute of Public Health's portal. Request by email.
Indicadores Clave Sistema Nacional de Salud (España)	https://inclasns.sanidad.gob.es/main.html	Mortality rates, disease incidence, and healthcare resources.	Country and regional	Spain	Publicly accessible through the Spanish Ministry of Health's portal.
Swedish mammography cohort (SMC) 2019 lifestyle (Sweden)	https://snd.gu.se/sv/catalogue/dataset/ext0018-6	Dietary habits, lifestyle factors, and health outcomes.	Individual	Sweden	Requires an application submission through the Swedish National Data Service (SND).
Cohort of Swedish men (COSM) (Sweden)	https://snd.gu.se/sv/catalogue/dataset/ext0015-4	Diet, lifestyle factors, and health outcomes.	Individual	Sweden	Requires an application through the Swedish National Data Service (SND).
Women's Lifestyle and Health (WLH) - 2003 (Sweden)	https://snd.gu.se/sv/catalogue/dataset/ext0153-2	Lifestyle factors, reproductive history, and health outcomes.	Individual	Sweden	Requires an application through the Swedish National Data Service (SND).

Malmö Diet Cancer (Sweden)	https://snd.gu.se/sv/catalogue/dataset/ext0012-1	Dietary habits, lifestyle factors, and cancer incidence.	Individual	Sweden	Requires an application through the Swedish National Data Service (SND).
Family Studies of Childhood Cancer - Genealogy Database (Sweden)	https://snd.gu.se/sv/catalogue/dataset/ext0123-1	Family histories, genetic information, and cancer incidence.	Individual	Sweden	Requires an application through the Swedish National Data Service (SND).
Slovenia and cancer - SLORA	http://www.slora.si/	Cancer incidence, mortality, and survival statistics.	Country and regional	Slovenia	Publicly accessible through the SLORA portal.
EUROpean CAncer REgistry based study on survival and care of cancer patients (EUROCARE)	https://www.iss.it/en/eurocare-il-progetto	Cancer survival rates, incidence, and patient demographics.	Country and regional	Europe	A request must be submitted to the EUROCARE project coordinators.
Spanish National Health Survey	https://www.sanidad.gob.es/estadEstudios/estadisticas/sisInfSanSNS/tablasEstadisticas/InfAnSNS.htm	Health status, lifestyle factors, and healthcare utilisation.	Country	Spain	Data is available through the Spanish Ministry of Health's portal.
Genomic Data Commons (GDC)	https://portal.gdc.cancer.gov/exploration/	Genomic sequences, clinical data, and biospecimen information.	Individual	Worldwide	Requires an application submission and approval.
Federal Health Reporting (Germany)	gbe-bund.de	Health statistics, disease prevalence, and healthcare services.	Country and regional	Germany	Publicly accessible through the Federal Health Reporting portal.
Wissenschaftliches Institut der AOK: WIdO (Germany)	https://www.wido.de/	Healthcare utilisation, prescription data, disease prevalence, and health outcomes.	Country and regional	Germany	Requires formal requests and approval from WIdO.
European Health for All database (HFA-DB)	https://gateway.euro.who.int/en/datasets/european-health-for-all-database/	Health indicators including mortality rates, morbidity, healthcare resources, and risk factors.	Country	Europe	Publicly accessible through the WHO European Health Information Gateway.
Dutch Lung Cancer Audit Database (DLCA)	https://dica.nl/dlca/home	Demographics, tumour characteristics, treatment details, and	Individual	Netherlands	Access requires formal application and approval from the Dutch Institute for Clinical Auditing (DICA).

		outcomes related to lung cancer.			
National Cancer Registry Ireland (NCR)	https://www.ncri.ie/data/data-we-collect	Cancer incidence, demographics, tumour characteristics, and survival data.	Individual	Ireland	Requires a request through a formal application to NCRI.
Dutch Surgical Colorectal Audit (DSCA)	https://dica.nl/dcra/documenten	Demographics, surgical procedures, postoperative outcomes, and quality indicators for colorectal surgery.	Individual	Netherlands	Requires formal application and approval from DICA.
Dutch Lung Surgery Audit (DLSA)	https://dica.nl/dlca/onderzoek	Demographics, surgical details, postoperative outcomes, and quality indicators for lung surgery.	Individual	Netherlands	Requires formal application and approval from DICA.
The Centre for Cancer Registry Data (ZfKD)	https://www.krebsdaten.de/Krebs/EN/Database/databasequery_step1_node.html	Cancer incidence, mortality, survival rates, and demographics.	Country and regional	Germany	Publicly accessible through the ZfKD database query system.
Inserm CépiDc (France)	https://opendata-cepidc.inserm.fr/	Mortality data, including causes of death and demographic information.	Country and regional	France	Publicly accessible through the Inserm CépiDc open data portal.
StatFin database (Finland)	https://pxdata.stat.fi/PxWeb/pxweb/en/StatFin/	Health, economy, and demographics.	Country and regional	Finland	Publicly accessible through Statistics Finland's portal.
National Board of Health and Welfare's Cancer Register	https://www.socialstyrelsen.se/en/statistics-and-data/statistics/statistical-databases/	Cancer incidence, demographics, tumour characteristics, and survival data.	Individual	Sweden	Requires a request through a formal application to the National Board of Health and Welfare.
Instituto Nacional de Estadística (INE)	https://www.ine.es/index.htm	Health, demographics, economy, quality of life	Country and regional	Spain	Publicly accessible through the INE portal.

Netherlands Cancer Registry (NCR)	https://iknl.nl/	Cancer incidence, demographics, tumour characteristics, and survival data.	Individual	Netherlands	Requires a request through a formal application to the NCR.
Epidemiological Strategy and Medical Economics (ESME)	https://recherche.unicancer.fr/en/programs/esme/	Clinical data, treatment patterns, outcomes, and healthcare resource utilisation for specific cancers.	Individual and Country	France	Requires formal application and approval from Unicancer.
EUSOMA database	https://www.eusoma.org/en/about%2dus/1-105-1-	Breast cancer diagnosis, treatment, outcomes, and quality indicators.	Individual	Europe	Requires formal application and approval from EUSOMA.
Statistik Austria	https://www.statistik.at/statistiken/bevoelkerung-und-soziales/gesundheit/krebserkrankungen	Health statistics, including disease prevalence, mortality rates, and healthcare services.	Country and regional	Austria	Publicly accessible through the Statistik Austria portal.
Statistics Poland	https://dbw.stat.gov.pl/en	Health, demographics, and economy.	Country and regionaç	Poland	Publicly accessible through the Statistics Poland portal.
WHO mortality database	https://www.who.int/data/data-collection-tools/who-mortality-database	Mortality data by cause of death, age, sex, and country.	Country	Worldwide	Publicly accessible through the WHO Mortality Database portal.
Global Cancer Observatory: Cancer Over Time WHO	https://gco.iarc.fr/en	Cancer incidence, mortality, prevalence, and survival statistics over time.	Country	Worldwide	Publicly accessible through the GCO portal.
Global Cancer Observatory: Cancer Today WHO	https://gco.iarc.fr/en	Current estimates of cancer incidence, mortality, and prevalence.	Country	Worldwide	Publicly accessible through the GCO portal.
Global Cancer Observatory: Cancer Tomorrow WHO	https://gco.iarc.fr/en	Projected cancer incidence and mortality estimates.	Country	Worldwide	Publicly accessible through the GCO portal.

Global Cancer Observatory: Cancer Causes WHO	https://gco.iarc.fr/en	Cancer risk factors and attributable fractions.	Country	Worldwide	Publicly accessible through the GCO portal.
NORDCAN Data	https://nordcan.iarc.fr/en/dataviz	Cancer incidence, mortality, prevalence, and survival statistics.	Country	Nordic countries	Publicly accessible through the NORDCAN portal.
ECIS - European Cancer Information System	https://ecis.jrc.ec.europa.eu/	Cancer incidence, mortality, survival rates, and prevalence estimates.	Country	Europe	Publicly accessible through the ECIS portal.
The Danish Diet, Cancer and Health cohort	https://www.cancer.dk/danish-cancer-institute/research-groups/diet-cancer-and-health/	Dietary habits, lifestyle factors, biological markers, and health outcomes.	Individual	Denmark	Requires formal application and approval from the Danish Cancer Society.
Bavaria Oncological Data Set	https://basisdatensatz.de/	Cancer diagnoses, treatments, outcomes, and patient demographics.	Individual	Bavaria region	Requires formal application and approval from the Bavarian Cancer Registry.
Cancer Incidence in Five Continents (CI5): WHO	https://ci5.iarc.fr/	Cancer incidence data from population-based cancer registries worldwide.	Country and regional	Worldwide	Publicly accessible through the CI5 portal.
EU statistics on income and living conditions (EU-SILC)	https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions	Income, poverty, social exclusion, and living conditions.	Household and individual	Europe	Requires formal application and approval from Eurostat.
Lithuanian Cancer Registry (NVI)	https://data.gov.lt/datasets/1437/	Cancer incidence and mortality data, demographics, tumour characteristics, and survival statistics.	Country	Lithuanian	Publicly accessible through the Lithuanian Open Data Portal.
Healthy Ireland Survey	https://www.ucd.ie/issda/	Health behaviours, lifestyle factors, and health status indicators, including smoking habits, alcohol	Individual	Ireland	Data is available through the Irish Social Science Data Archive (ISSDA). Requires a data request form and agreement to the terms of use.

		consumption, physical activity, diet, and mental health.			
Irish Health Survey	https://www.ucd.ie/issda/	Chronic conditions, health care utilisation, and preventive health behaviours.	Individual	Ireland	Available through the Irish Social Science Data Archive (ISSDA). Requires submission of a data request form and adherence to specific usage guidelines.

Version History

Version	Description	Date completed
v1.0	First draft	05/11/2024
v2.0	iBeChange revision	29/11/2024