REGISTERED REPORT STAGE 1: STUDY DESIGN

Brain and Behavior Open Access WILEY

Psychosocial factors and cancer incidence (PSY-CA): Protocol for individual participant data meta-analyses

Lonneke A. van Tuijl[1] | Adri C. Voogd[2,3,4] | Alexander de Graeff[5] | Adriaan W. Hoogendoorn[6,7] | Adelita V. Ranchor[1] | Kuan-Yu Pan[6] | Maartje Basten[8] | Femke Lamers[6] | Mirjam I. Geerlings[8] | Jessica G. Abell[9] | Philip Awadalla[10,11,12] | Marije F. Bakker[8] | Aartjan T. F. Beekman[6] | Ottar Bjerkeset[13,14] | Andy Boyd[15] | Yunsong Cui[16] | Henrike Galenkamp[17] | Bert Garssen[1] | Sean Hellingman[18] | Martijn Huisman[19,20] | Anke Huss[21] | Melanie R. Keats[22] | Almar A. L. Kok[6,19] | Annemarie I. Luik[23,24] | Nolwenn Noisel[25] | N. Charlotte Onland-Moret[8] | Yves Payette[25] | Brenda W. J. H. Penninx[6] | Lützen Portengen[21] | Ina Rissanen[8] | Annelieke M. Roest[26] | Judith G. M. Rosmalen[27] | Rikje Ruiter[23,28] | Robert A. Schoevers[29] | David M. Soave[18,10] | Mandy Spaan[30] | Andrew Steptoe[9] | Karien Stronks[17] | Erik R. Sund[13,31,32] | Ellen Sweeney[16] | Alison Teyhan[15] | Ilonca Vaartjes[8] | Kimberly D. van der Willik[23,30] | Flora E. van Leeuwen[30] | Rutger van Petersen[8] | W. M. Monique Verschuren[8,33] | Frank Visseren[34] | Roel Vermeulen[21] | Joost Dekker[35,36]

[1] Department of Internal Medicine, Maasstad Hospital, Rotterdam, The Netherlands

[2] Department of Internal Medicine, Division of Medical Oncology, GROW, Maastricht University Medical Centre, Maastricht, The Netherlands

[3] Department of Epidemiology, GROW, Maastricht University, Maastricht, The Netherlands

[4] Department of Research, Netherlands Comprehensive Cancer Organization (IKNL), Utrecht, The Netherlands

[5] Department of Medical Oncology, Cancer Center University Medical Center, University of Utrecht, Utrecht, The Netherlands

[6] Amsterdam UMC, Department of Psychiatry, Amsterdam Public Health Research Institute, Vrije Universiteit, Amsterdam, The Netherlands

[7] GGZ inGeest Specialized Mental Health Care, Amsterdam, The Netherlands

[8] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht and Utrecht University, Utrecht, the Netherlands

[9] Department of Behavioural Science and Health, University College London, London, UK

[10] Ontario Institute for Cancer Research, Toronto, Ontario, Canada

[11] Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

[12] Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

[13] Faculty of Nursing and Health Sciences, Nord University, Bodø, Norway

[14] Faculty of Medicine and Health Sciences, Department of Mental Health, Norwegian University of Science and Technology, Trondheim, Norway

[15] Bristol Medical School, Population Health Sciences, University of Bristol, Bristol, UK

[16] Atlantic Partnership for Tomorrow's Health, Faculty of Health, Dalhousie University, Halifax, Nova Scotia, Canada

[17] Department of Public and Occupational Health, Amsterdam UMC, and Amsterdam Public Health Research Institute, University of Amsterdam, Amsterdam, Netherlands

[18] Department of Mathematics, Wilfrid Laurier University, Waterloo, Canada

[19] Amsterdam UMC, Department of Epidemiology & Data Science, Amsterdam Public Health institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[20] Department of Sociology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[21] Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands

[22] School of Health and Human Performance, Faculty of Health, Dalhousie University, Halifax, Canada

[23] Department of Epidemiology, Erasmus MC–University Medical Center, Rotterdam, The Netherlands

[24] Department of Child and Adolescent Psychiatry/Psychology, Erasmus MC–University Medical Center, Rotterdam, The Netherlands

[25] CARTaGENE, CHU Sainte-Justine, 3175, Chemin de la Côte-Sainte-Catherine, Montréal, Québec, Canada

[26] Department of Developmental Psychology, University of Groningen, Groningen, The Netherlands

[27] Departments of Psychiatry and Internal Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[28] Department of Internal Medicine, Maasstad, Rotterdam, The Netherlands

[29] Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

[30] Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[31] Department of Public Health and Nursing, HUNT Research Centre, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

[32] Levanger hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway

[33] Centre for Nutrition, Prevention and Health Services, National Institute for Public Health and the Environment, Utrecht, the Netherlands

[34] Department of Vascular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands

[35] Amsterdam Public Health Research Institute, Amsterdam, Noord-Holland, The Netherlands

[36] Department of Rehabilitation Medicine and Department of Psychiatry, Amsterdam UMC - VUMC, Amsterdam, Noord-Holland, The Netherlands

Correspondence
Lonneke A. van Tuijl, Health Psychology Section, HPC FA12, University Medical Center Groningen, Hanzeplein 1, Postbus 30 001, 970 RB Groningen, The Netherlands.
Email: L.a.van.tuijl@umcg.nl

Abstract

Objectives: Psychosocial factors have been hypothesized to increase the risk of cancer. This study aims (1) to test whether psychosocial factors (depression, anxiety, recent loss events, subjective social support, relationship status, general distress, and neuroticism) are associated with the incidence of any cancer (any, breast, lung, prostate, colorectal, smoking-related, and alcohol-related); (2) to test the interaction between psychosocial factors and factors related to cancer risk (smoking, alcohol use, weight, physical activity, sedentary behavior, sleep, age, sex, education, hormone replacement therapy, and menopausal status) with regard to the incidence of cancer; and (3) to test the mediating role of health behaviors (smoking, alcohol use, weight, physical activity, sedentary behavior, and sleep) in the relationship between psychosocial factors and the incidence of cancer.

Methods: The psychosocial factors and cancer incidence (PSY-CA) consortium was established involving experts in the field of (psycho-)oncology, methodology, and epidemiology. Using data collected in 18 cohorts ($N = 617,355$), a preplanned two-stage individual participant data (IPD) meta-analysis is proposed. Standardized analyses will be conducted on harmonized datasets for each cohort (stage 1), and meta-analyses will be performed on the risk estimates (stage 2).

Conclusion: PSY-CA aims to elucidate the relationship between psychosocial factors and cancer risk by addressing several shortcomings of prior meta-analyses.

KEYWORDS
cancer risk, depression, health behaviors, meta-analysis, psycho-oncology

1 | INTRODUCTION

Psychosocial factors such as depression, general distress, and low social support have long been theorized to increase cancer risk (Dalton et al., 2002). Findings from prior research studying the association between psychosocial factors and cancer are mixed. Two meta-analyses focusing on depression concluded that there was a small, potentially trivial, effect on cancer risk (McGee et al., 1994; Oerlemans et al., 2007). Another meta-analysis of the published literature indicated that depression (combined hazards ratio [HR] = 1.29), psychosocial factors relating to stress-prone personality or poor coping style (combined HR = 1.08), and psychosocial factors relating to emotional distress or poor quality of life (combined HR = 1.13) increased the risk for all cancer outcomes and, when collapsing psychosocial factors across subtypes, especially for lung cancer (combined HR = 1.23) (Chida et al., 2008). However, the included studies varied greatly in the psychosocial factors investigated and the cancer endpoint of interest. It is crucial to use clearly and specifically defined psychosocial factors as they can lead to distinct physiological and behavioral effects (O'Donovan et al., 2010). These effects may increase risk for specific cancers given their unique etiologies. Furthermore, published studies vary greatly in the confounders adjusted for (if any), making reliability and interpretation of outcome debatable. Rather than including studies where analyses have been determined by the original authors, two-stage individual participant data (IPD) meta-analysis refers to the (re-)analysis of original data for each cohort using a standardized approach (stage 1) before combining in a meta-analysis (stage 2) (Tierney et al., 2020). IPD meta-analyses of cohorts have the potential to produce more reliable results than meta-analyses of published findings (Stewart & Parmar, 1993) as one can ensure a consistent definition of the psychosocial factors, specific cancer endpoints, and key confounders adjusted for across all included cohorts.

Evidence remains limited regarding how psychosocial factors increase cancer risk. Theory postulates several possible, potentially interrelated, pathways that link psychosocial factors and cancer, including angiogenesis, endocrine mechanisms, immunosuppression, impairments in DNA repair, and inflammation (Lutgendorf et al., 2007). While health behaviors as a potential pathway between psychosocial factors and cancer have received little attention, they deserve consideration given the established relationship between health behaviors and psychosocial factors (Strine et al., 2008; Verger et al., 2009), and between health behaviors and cancer (Biswas et al., 2015; Chen et al., 2018; Kerr et al., 2017). To date, studies have most often considered health behaviors as confounders, rather than playing a direct role in the association between psychosocial factors and cancer risk. If health behaviors explain the relationship between psychosocial factors and cancer risk, this may justify offering health behavior interventions in at-risk groups such as individuals who are depressed and smoke.

The effects of psychosocial factors on cancer development may depend on the presence or absence of health behaviors, somatic factors or demographic factors. If both psychosocial factors and health behaviors or somatic factors play a causal role in cancer development, they may interact with each other. If there is interaction, the presence of both factors puts a person at a higher risk for cancer than what would be expected based on the sum of the risk of each factor alone. This may be the case when psychosocial factors and health behaviors or somatic factors affect cancer development via the same or interrelated pathways. Factors which have been related to cancer risk and may interact with psychosocial factors include smoking (Knekt et al., 1998), weight (Kerr et al., 2017), alcohol use (Pelucchi et al., 2011), physical activity (Kerr et al., 2017), sedentary behavior (Kerr et al., 2017), sleep duration and sleep quality (Hurley et al., 2015), menopausal status (Trichopoulos et al., 1972), hormone replacement therapy (Vecchia et al., 2001), age (Thakkar et al., 2014), sex (White et al., 2018), and education level (Mouw et al., 2008). For example, in one study, depressive symptoms increased the risk of colorectal cancer particularly in overweight women (Kroenke et al., 2005), and another study found that the effect of depressive symptoms on cancer risk was increased at higher levels of cigarette smoking (Linkins & Comstock, 1990). Studying interactions provides insight into the mechanisms leading to cancer development and also shows for which subgroups the association between psychosocial factors and cancer incidence is most prominent and thus could benefit most from preventive interventions.

Health behaviors may not only interact with psychosocial factors, but may also function as mediators situated in the pathway from psychosocial factors to the development of cancer. Symptoms of depression, for example, have been linked to smoking initiation and the amount of smoking (Steuber & Danner, 2006), increased alcohol use (Bulloch et al., 2012), weight gain, weight loss, obesity (Blaine, 2008), decreased physical activity (Roshanaei-Moghaddam et al., 2009), increased sedentary behavior (Roshanaei-Moghaddam et al., 2009), and sleep disturbances (Benca et al., 1997), all of which have subsequently been associated with an increased cancer risk (Biswas et al., 2015; Chen et al., 2018; Kerr et al., 2017). While weight is not a health behavior, we refer to this as a health behavior given the association with several other health behaviors, specifically diet and physical activity. Despite numerous allusions to the potential mediating role of health behaviors in the relationship between psychosocial factors and cancer (Chida et al., 2008; Dalton et al., 2002), there are remarkably few studies in which this has been tested.

The psychosocial factors and cancer incidence (PSY-CA) consortium was established to investigate whether a variety of psychosocial factors increase the risk of cancer. The investigated psychosocial factors include diagnosed depressive disorder and depressive symptoms (Jia et al., 2017) (here forth referred to as depression), diagnosed anxiety disorder and anxiety symptoms (Chen et al., 2018) (here forth referred to as anxiety), (recent) loss events (Dalton et al., 2002), perceived low social support (Idahl et al., 2018), relationship status (Randi et al., 2004), general distress (Peled et al., 2008), and neuroticism (Schapiro et al., 2001). The cancer endpoints include any cancer and the four most prevalent cancers worldwide (excluding nonmelanoma skin cancer): breast cancer, lung cancer, prostate cancer, and colorectal cancer. We also categorize cancers for which common causal factors are known, namely: smoking-related cancers and alcohol-related cancers, as psychosocial factors may increase smoking and alcohol use (Strine
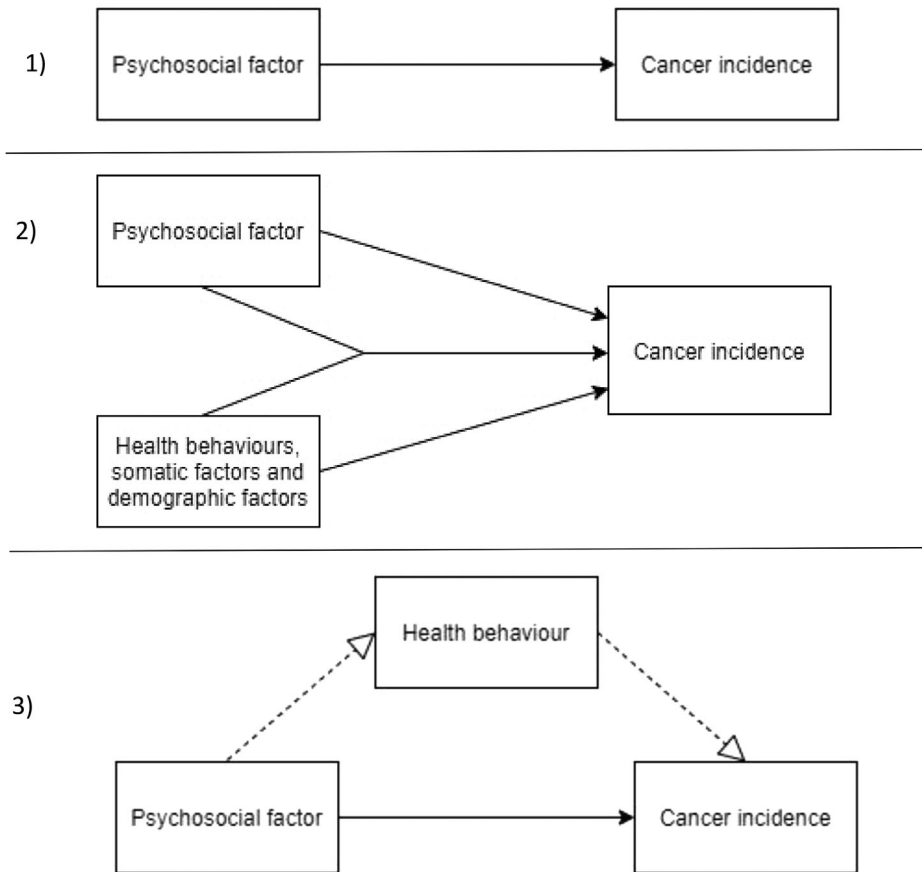
FIGURE 1 The three models that are researched in psychosocial factors and cancer incidence (PSY-CA): (1) the longitudinal association between psychosocial factors and cancer incidence, (2) the interaction between psychosocial factors and health behaviors, somatic factors, and demographic factors in cancer incidence, and (3) the mediating role of health behaviors in the longitudinal association between psychosocial factors and cancer incidence

et al., 2008; Verger et al., 2009), thereby increasing risk for these cancers. The goals of the PSY-CA consortium are (1) to test whether psychosocial factors (depression, anxiety, recent loss events, subjective social support, relationship status, general distress, and neuroticism) are associated with the incidence of any cancer (breast, lung, prostate, colorectal, smoking-related cancer, and alcohol-related cancer); (2) to test the interaction between psychosocial factors and factors related to cancer risk (smoking, alcohol use, weight, physical activity, sedentary behavior, sleep, age, sex, education, hormone replacement therapy, and menopausal status) in the incidence of cancer; (3) to test the mediating role of health behaviors (smoking, alcohol use, weight, physical activity, sedentary behavior, and sleep) in the relationship between psychosocial factors and the incidence of cancer (see Figure 1). Specific hypotheses have been formulated (Appendix 1).

2 | METHODS

2.1 | Design overview

Preplanned two-stage IPD meta-analyses are performed. We apply the Maelstrom guidelines (Fortier et al., 2016) to create harmonized variables across the 18 cohorts. Data are analyzed in each cohort (stage 1) and the outputs are used in a meta-analysis (stage 2).

2.2 | PSY-CA consortium

The consortium consists of the steering group (LvT, JD, AV, AdG, and AVR), three main researchers (LvT, MB, and K-YP), representatives from each participating cohort, and selected experts in the field of psycho-oncology, epidemiology, methodology, and statistics. Meetings are held at least two times a year with the first formal consortium meeting having taken place on March 2019. During the meetings, consensus is reached on the objectives, approach, and interpretation of findings. The project leader (JD), the steering group, and the representatives of the Dutch cohorts are responsible for the formal management of the study.

2.3 | Preregistration

The PSY-CA study has been preregistered in PROSPERO:

2.4 | Ethics

The ethics approval for PSY-CA was waived by the Medical Ethics Review Committee of VU University Medical Center (2018.101). For inclusion in PSY-CA, ethics approval was granted for each study by the local institution or through appropriate national research governance frameworks.

2.5 | Inclusion/exclusion criteria

2.5.1 | Cohorts

Cohorts were eligible to take part in PSY-CA if following criteria were met:

1. a valid and reliable measure of depression, anxiety, recent loss events, social support, general distress, and/or neuroticism;
2. availability of an objective measure of cancer diagnosis during follow-up or the potential to get this information through, for example, linkage with a cancer registry;
3. availability of data regarding smoking, alcohol, sex, and age; and
4. a prospective study design (i.e., psychosocial factors were measured before cancer incidence).

Cohorts were not eligible if there was no information about a history of cancer at baseline. Initially, relatively objective social support (i.e., social network size) and hopelessness were concepts also included in the first criterion, however as most cohorts did not have a measure of this, these concepts were subsequently dropped. Objective social support was replaced with relationship status. One cohort—Prospect-EPIC (see Table 1)—initially appeared to have information about depression and anxiety diagnoses through a psychiatric registry. However, on closer inspection, this data appeared to be incomplete. As relationship status was measured in Prospect-EPIC, it remained included in the study.

2.5.2 | Participants

Across all analyses, participants were excluded if they had a cancer diagnosis (based on [cancer] registry data or self-report) at baseline or in the past (including in situ carcinomas and neoplasms of undetermined behavior [i.e., benign/malignant status undetermined]), with the exception of nonmelanoma skin cancer. Participants who had refused linkage with an external registry were also excluded from any analysis. People with a cancer diagnosis within one year from baseline were excluded from the analysis.

2.6 | Search strategy and eligible studies

In preparation of PSY-CA, a feasibility study was conducted to identify potential cohorts (December 2015 toMarch 2017). An extensive search for all relevant Dutch cohorts was conducted using the network of experts participating in PSY-CA. The coordinators of these cohorts were approached and invited to take part in PSY-CA. In order to increase the number of cohorts, international cohorts that fulfilled the inclusion criteria were identified through the BioShare consortium (which is now linked to the Public Population Project in Genomics and Society) (http://www.p3gconsortium.org/about-p3g), Integrative Analysis of Longitudinal Studies of Aging network (www.ialsa.org/). Consortium members were also asked if they knew of the existence of any international cohorts that met the inclusion criteria. In addition, a literature review was conducted in preparation of the feasibility study.

During the feasibility study, coordinators of the candidate cohorts were contacted to check whether the inclusion and exclusion criteria were met, and to outline any other potential issues related to costs or ethical issues, for example. The included cohorts (11 cohorts in the Netherlands and seven cohorts in the United Kingdom, Norway, and Canada) are outlined in Table 1.

PSY-CA is set up in such a way that after the project has finished, additional cohorts can be incorporated by applying the harmonization manual (outlined in data handling below) to the data and running the standardized analyses scripts.

2.7 | Variables

2.7.1 | Psychosocial factors

The relationship between the following seven psychosocial factors and cancer incidence is analyzed: depressive symptoms or clinical depression (i.e., major depressive disorder, dysthymia), anxiety symptoms or anxiety disorders (excluding specific phobias), recent loss events (defined specifically as the loss of an immediate family member or partner in the last 12 months), perceived social support, relationship status, general distress (specified as scored on the Mental Health Inventory of the Short-form health survey (SF-36) or the RAND36 (Ware & Sherbourne, 1992)), and neuroticism. Only validated or previously published measures of the psychosocial factors are used.

2.7.2 | Cancer

The primary outcomes in PSY-CA are incidence of all cancers, smoking-and alcohol-related cancers, and the four most prevalent cancers in the Netherlands: breast cancer (females only), colorectal cancer, lung cancer, and prostate cancer (males only). Cancers were classified as

TABLE 1 Overview cohorts participating in the psychosocial factors and cancer incidence (PSY-CA) study

Name of Cohort	Number of subcohorts included[b]	Organizations (country)	Number of participants[a]	Approx. max follow-up duration (years) for cancer diagnosis	Reference
Ontario Health Study	1	University of Toronto (Canada)	163,257	10	(Borugian et al., 2010; Dummer et al., 2018)
Lifelines	1	University Medical Center Groningen (The Netherlands)	152,000	13	(Scholtens et al., 2015)
Nord-Trøndelag Health Study (HUNT)	2	Norwegian University of Science and Technology (Norway)	62,237	13–24	(Krokstad et al., 2013)
CARTaGENE	1	Centre Hospitalier Universitaire Sainte Justine (Canada)	43,000	10	(Awadalla et al., 2013; Borugian et al., 2010; Dummer et al., 2018)
Atlantic PATH	1	Dalhousie University (Canada)	34,169	10	(Borugian et al. 2010; Sweeney et al., 2017)
European Prospective Investigation into Cancer and Nutrition (MORGEN-EPIC)	1	National Institute for Public Health and the Environment (RIVM) (The Netherlands)	23,100	24	(Beulens et al., 2010; Riboli, 1992, 2002)
Healthy Life in an Urban Setting (HELIUS)	1	Amsterdam University Medical Centers and Amsterdam Municipal Health Service (The Netherlands)	19,932	8	(Snijder et al., 2017; Stronks et al., 2013)
European Prospective Investigation into Cancer and Nutrition (Prospect-EPIC)	1	University Medical Center Utrecht (The Netherlands)	17,357	24	(Beulens et al., 2010; Boker et al., 2001; Riboli, 1992, 2002)
Dutch occupational and Environmental Health Cohort Study (AMIGO)	1	Utrecht University (The Netherlands)	14,829	5	(Slottje et al., 2014)
Avon Longitudinal Study of Parents and Children (ALSPAC)	1	University of Bristol (England)	14,541	20	(Fraser et al., 2013)
Second Manifestations of ARTerial disease (SMART)	1	University Medical Center Utrecht (The Netherlands)	11,881	12	(Simons et al., 1999)
Rotterdam Study	3	Erasmus MC University Medical Center (The Netherlands)	11,740	8–14	(Hofman et al., 2015)
English Longitudinal Study of Ageing (ELSA)	1	University College London (England)	11,391	14	(Steptoe et al., 2013)
Whitehall-II study (WH-II)	1	University College London (England)	10,308	11	(Marmot & Brunner, 2005)
OMEGA-II	1	The Netherlands Cancer Institute (The Netherlands)	10,000	8	(van den Belt-Dusebout et al., 2016; Spaan et al., 2016)
Utrecht Health Project (UHP)	2	University Medical Center Utrecht (The Netherlands)	10,000	11–19	(Grobbee et al., 2005)

(Continues)

TABLE 1　(Continued)

Name of Cohort	Number of subcohorts included[b]	Organizations (country)	Number of participants[a]	Approx. max follow-up duration (years) for cancer diagnosis	Reference
Longitudinal Aging Study Amsterdam (LASA)	1	Amsterdam University Medical Centers (The Netherlands)	4632	28	(Hoogendijk et al., 2016; Huisman et al., 2011)
Netherlands Study of Depression and Anxiety (NESDA)	1	Amsterdam University Medical Centers, on behalf of the NESDA consortium (www.nesda.nl) (The Netherlands)	2981	15	(Penninx et al., 2008)

Note: In some cohorts a measurement wave other than baseline is used in PSY-CA due to the absence of a measure relating to one of the psychosocial factors outlined in the hypotheses.
[a]This is before applying any exclusion criteria (e.g., a history of cancer) and based on baseline adult sample sizes.
[b]Subcohorts are limited to those that are treated as subcohorts in the meta-analyses. For certain cohorts, subcohorts were combined where subsample sizes were too small otherwise (i.e., <1000) and combining resulted in minimal or no loss of data.

TABLE 2　Overview cancer sites considered smoking-related and/or alcohol-related (ICD10 codes)

Both smoking- and alcohol- related cancer sites	Smoking-related cancer sites	Alcohol-related cancer sites
Tongue (C01)	Nasopharynx (C11)	Liver and intrahepatic bile ducts (C22)
Other and unspecified parts of the tongue (C02)	Stomach (C16)	Breast (C50)
Gum (C03)	Liver and intrahepatic bile ducts (C22)	
Floor of mouth (C04)	Pancreas (C25)	
Palate (C05)	Nasal cavity (C30, excluding C30.1—middle ear)	
Other and unspecified parts of mouth (C06)	Accessory sinuses (C31)	
Tonsil (C09)	Bronchus and lung (C34)	
Oropharynx (C10)	Cervix uteri (C53)	
Piriform sinus (C12)	Ovary (C56)	
Hypopharynx (C13)	Kidney (C64)	
Other and ill-defined sites in lip, oral cavity and pharynx (C14)	Renal pelvis (C65)	
Oesophagus (C15)	Ureter (C66)	
Colorectal (C19-C20)	Bladder (C67)	
Larynx (C32)	Myeloid leukemia (C92)	

smoking- or alcohol-related as listed by the International Agency for Research on Cancer (International Agency for Research on Cancer, 2019) and double-checked by the medical oncologist in the PSY-CA steering group (AdG; see Table 2). Cancer site is determined with codes from the International Classification of Diseases (ICD) version 10 or ICD-Oncology version 3 codes for the majority of cohorts, with few based on ICD version 9 codes. Only the first cancer diagnosis during follow-up is considered, and (where available) in situ carcinomas and neoplasms of undetermined behavior are included as the latter could be malignant and the former may develop into cancer later if left untreated. Analyses are also done excluding those with carcinoma

in situ and cancers with undetermined behavior to explore whether conclusions hold. All studies determine cancer diagnoses through linkage with national cancer registries, with the exception of CARTaGENE and Rotterdam Study. In both these cohorts, other registries or databases are used to supplement missing information from cancer registries including hospital visits, insurance claims, and GP records. The date of cancer diagnosis is considered as the date of cancer incidence. Where cohorts only provided month and year, a fixed date (15th) is applied for day of diagnosis of cancer. Where cohorts only provided year of diagnosis, June 30th is assumed to be the date of diagnosis.

2.8 | Harmonization and data cleaning

Data harmonization ensures the quality of the results, and the interpretability. Harmonization of data across cohorts also enables the use of standardized scripts for stage 1 analyses (see statistical analysis below) requiring minimal user input. We apply the Maelstrom guidelines (Fortier et al., 2016) to create individual harmonization manuals for each cohort providing guidance on how to recode data to create the variables required for PSY-CA. Definitions of the variables to be derived are agreed upon within the consortium. Local researchers at each cohort harmonize the data and receive a script to run a number of basic checks (e.g., checking for proportion of missing data). The basic checks are then reviewed by two researchers (LvT and MB) as an additional check of adherence to the manuals.

2.8.1 | Missing data

Previously defined, cohort-specific approaches to dealing with missing data is applied within a given measure (e.g., questionnaire). Where no such approach has previously been defined, the general rule applied is to substitute person-mean for up to 20% missing responses for a measure. This rule is based on previous studies comparing ways to deal with item-level missing data, specifically in measures of depression (Bono et al., 2007; Shrive et al., 2006). Missing responses or responses equivalent to "I don't know" are coded as missing data. The only exception is the family history of cancer variables where "I don't know" is coded as no family history of (the specific) cancer.

2.8.2 | Unlikely values and extreme outliers

Local researchers harmonizing the cohort data are instructed to investigate extreme, unlikely values and recode these to missing if there is sufficient support that these are errors (e.g., a very high BMI that is markedly higher than the BMI reported at a follow-up wave for a given participant).

Across all cohorts, extreme outliers are defined as values that are more than three times the interquartile range above the third quartile or below the first quartile, and are truncated to the cut-offs, respectively. The exception to this rule is variables that contain true zeroes as the lowest possible score as these are likely to be skewed (e.g., pack years where all never smokers score zero). For these variables, only the upper extreme values are truncated. The number of cases that are capped, and the replacement value are recorded for all cohorts and double-checked.

2.9 | Statistical analysis

PSY-CA employs a two-stage design. In stage 1, local researchers at the cohort level run a standardized R script over the harmonized dataset, and subsequently provide all output generated to the main researchers

(LvT, MB, and K-YP). In stage 2, the output from stage 1 across all cohorts is pooled in the meta-analyses. As such, the main researchers do not have direct access to cohort raw data. However, in the event that further clarification is required from specific cohort data, subsequent scripts are sent to the local researchers to gain additional information.

2.9.1 | Stage 1

For the analyses related to question one (relationship between psychosocial factors and cancer) and question two (interaction), Cox regression models are used. For question three (mediation), different regression models (logistic and multiple) are used to test the path between the psychosocial factor and the mediator, dependent on whether the mediator is categorical or continuous. Cox regression models are used for the path between the psychosocial factor and cancer, and between the mediator and cancer.

Across all research questions, entry age is the age at baseline, while exit age is the age at cancer incidence, death, or end of cancer follow-up period of the respective cohort (whichever comes first). Note that several cohorts are ongoing but, for the purposes of PSY-CA, are capped to the moment of linkage with the cancer or vitality registry (whichever comes first). Where another type of cancer occurs (e.g., lung cancer) after cancer endpoint being modeled (e.g., breast cancer), participants are censored at the age of first diagnosis (Ji et al., 2020).

For the first two research questions, the following models are run:

> Model 1: univariable—which includes the year of birth and the psychosocial factor.
>
> Model 2: minimum—which additionally adjust for: education (high, low, with mid-level as reference category) and country of birth (i.e., whether the participant and [where information available] his/her parents were born in the country where the cohort is measured). These confounders are available in all studies with the exception of education in one subcohort of the HUNT cohort where occupation is used instead (high-ranking profession, low-ranking profession, with mid-level as a reference category). Furthermore, in this model we stratify the baseline rates on sex (with the exception of three all-female studies, and sex-specific cancer endpoints).
>
> Model 3: maximum—which additionally adjust for, where available, the following measured at baseline:

- All cancer outcomes: current anti-depressant use, weekly alcohol intake, hours of physical activity per week (metabolic equivalent [MET], if available), body mass index (BMI), pack years, and smoking status (former, current, with never smokers as reference), family history of cancer (where information about family history of cancer endpoint are not available, e.g., family history of breast cancer where breast cancer is the endpoint).
- Breast cancer outcome: parity (distinguishing between three or more [full-term] pregnancies, one to two pregnancies, with zero pregnancies as the reference category), contraceptive use,

menopause status, age at menarche, and a family history of breast cancer.

- Colorectal cancer outcome: sedentary behavior and family history of colorectal cancer.
- Lung cancer outcome: a family history of lung cancer.
- Prostate cancer outcome: a family history of prostate cancer.

Covariates were included where there was no more than 40% missing. A number of the variables adjusted for are considered to potentially interact with psychosocial factors for the second research question, and as mediators for the third research question. Across all models, subgroup analyses are conducted by sex. Furthermore, maximum models are explored where only covariates with no more than 10% missing are included, and excluding education (which may overcorrect for the role of health behaviours), and the minimum model is rerun with the sample of the maximum model for comparability. All additional models, subgroup and sensitivity analyses (outlined below in specific research questions) are considered explorative.

2.9.2 | Specifics of research question one

Additional subgroup analyses are conducted based on: cancer stage at diagnosis (stages 1–2, stages 3–4), age group (younger [18–40], mid [41–64], and older [65+]). Additional sensitivity analysis is conducted where borderline or underdetermined cancers and carcinoma in situ are not considered cancer diagnosis (i.e., not an event), and these participants are censored at the moment of this diagnosis. Further sensitivity analysis is conducted where follow-up is capped to 5 years, 10 years, 15 years, and 20 years (where possible). When testing the role of the psychosocial factors other than depression, an additional model includes symptoms of depression or depression diagnosis (if available in the cohort) along with all the confounders listed above to explore the specificity of the psychosocial factor in the risk of cancer. Furthermore, additional explorative analyses include all psychosocial factors entered simultaneously in the model.

2.9.3 | Specifics of research question two

Cancer risk factors tested are based on smoking, alcohol use, weight, physical activity, sedentary behavior, sleep duration and sleep quality, age, sex, education, hormone replacement therapy, and menopausal status. Interaction is assessed by first entering the main effects of the psychosocial factor, the cancer risk factor and an interaction term between the psychosocial factor and cancer risk factor into the Cox models outlined above, to test for multiplicative interaction. Effect estimates of the psychosocial factor, cancer risk factor, and the interaction term are then used to calculate the relative excess risk due to interaction (RERI), a measure of additive interaction. Where interactions are significant, interpretation of the interaction effects is derived as follows: four categories are created (2 [high/low psychosocial factor] ×2 [high/low cancer risk factor]), and included in the Cox mod-

els outlined above. Further interpretation is derived by estimating the effect of psychosocial factors within subgroups of the cancer risk factor (high vs. low cancer risk factor) and estimating the effect of the cancer risk factor within subgroups based on psychosocial factor (high vs. low psychosocial factor). Additional subgroup analyses may be explored depending on the results from research question one.

2.9.4 | Specifics of research question three

Mediators tested are smoking, alcohol use, weight, physical activity, sedentary behavior, sleep duration, and sleep quality. Mediators are measured at baseline. Three paths are tested in the mediation analyses (for each mediator): path a (psychosocial factor to mediator), path b (mediator to cancer incidence), and path c (the direct path from psychosocial factor to cancer incidence while controlling the mediator). Path a is estimated with linear regression (for continuous mediators) and (multinomial) logistic regression (for categorical mediators). Cox regression models are used to estimate paths b and c. The indirect effect is the product of a × b, the direct effect is c', and the total effect is a × b + c. Additional explorative analyses will be conducted where all mediators are entered simultaneously in the models. Subgroup analysis by sex is conducted. Additional subgroup analysis will explore differences in cancer stage at diagnosis (stages 1–2, stages 3–4) specifically when looking at physical activity as a mediator in prostate cancer. Further explorative analyses may be performed based on the findings from research questions one and two.

2.9.5 | Stage 2 (meta-analysis)

Random-effects meta-analyses are performed. Cohorts are included in a given meta-analysis when there are at least 10 cancer events and the sample size of the cohort (subgroup) is at least 200. Leave-one-out analyses are conducted to identify influential cohorts. Cohorts are considered influential if, upon exclusion of the cohort, the between-study heteogeneity or effect size substantially changes.

Specifically, for question one, the hazards effects (and robust standard errors) of the psychosocial factors are meta-analyzed. For question two, meta-analyses are conducted on hazard effects of the psychosocial factor × cancer risk factor interaction terms (multiplicative interaction) and on the RERI estimates (additive interaction). In question three, overall estimates of all the paths and indirect effect are obtained by carrying out separate univariate meta-analyses with random effects.

Sensitivity analyses will be performed (where at least two studies with sufficient power are included in the meta-analysis) for cohorts that determine depression and anxiety through use of clinical interviews. Additional sensitivity analyses are conducted including only cohorts that are recruited from the general population. Finally, moderators of effect size are explored including when the cohort started, and whether the cohort took place in the Netherlands or not.

2.10 | Power analysis

To test the power of our IPD meta-analysis, we ran a power simulation study similar to that of Ensor et al. (2018), with a focus on depression. The study information that we used for the simulation study involved the total number of participants, the prevalence of depression at the baseline measurement of the study (estimated where this was not yet known), and the expected number of cancer cases of two types of cancer: lung cancer and any cancer (anticipated smallest and largest categories, respectively). Requiring 80% power, an alpha of 0.05 (two-sided testing), and using fixed-effects meta-analysis of HR from Cox regression models, our simulation study showed that regarding main effects the minimal detectable effect size for is HR = 1.04 for any cancer and HR = 1.12 for lung cancer.

Regarding the interaction analyses, our calculations (based on the general convention that the sample size in a single trial should be increased approximately four times to detect the interaction effect (Brookes et al., 2004; McClelland & Judd, 1993)) showed that the minimal detectable effect size for any cancer is HR = 1.08 for any cancer and HR = 1.25 for lung cancer. Regarding the mediation analyses, our calculations (based on an inflation factor of two, compared to testing the main effects) showed that the corresponding minimal detectable effect size is HR = 1.06 for any cancer and HR = 1.18 for lung cancer. The inflation factor of two for mediation analysis was found as an upper bound by comparing sample sizes needed for main effects with those for mediation effects using Baron and Kenny's test assuming different effect sizes (Fritz & MacKinnon, 2007).

2.11 | Interpretation

The hypotheses tested include four psychosocial factors (depression [symptoms], anxiety [symptoms], recent loss events, and perceived social support), four health behaviors (smoking, alcohol, physical activity, and weight), and seven cancer outcomes (see Appendix 1). It is important to specify that the interpretation of the results is done holistically, and not based on a single association (i.e., "cherry picking"). Through triangulation of the evidence from the different analyses, we conclude if there is statistical support of an association between psychosocial factors and cancer, and by extension whether there is evidence for interaction of or mediation by health behaviors. Interpretation is done by looking at the obtained HRs (or beta-coefficients) and 95% confidence intervals (CIs) and by exploring consistency and robustness of the findings. Additionally, the associations between a number of further psychosocial factors, other health and demographic factors, and cancer are studied. The results of these additional analyses are considered to be exploratory. Subgroup and sensitivity analyses are considered to be exploratory as well.

3 | DISCUSSION

Previous meta-analyses investigating the role of psychosocial factors in cancer incidence have shown mixed findings (Chida et al., 2008; McGee et al., 1994; Oerlemans et al., 2007). While this is partly explained by differences in the types of psychosocial factors and cancer endpoints, many studies in these meta-analyses pose further limitations, in particular the absence of adjustment for key confounders. PSY-CA aims to elucidate the relationship between psychosocial factors and cancer incidence by employing clearly defined psychosocial factors measured with reliable instruments. Through harmonization of the data across cohorts, strict definitions of the psychosocial factors are applied, thereby increasing interpretability. Furthermore, PSY-CA considers several different cancer endpoints in addition to any cancer endpoint, which is crucial given the distinct etiologies of cancers. By employing two-stage IPD meta-analyses, PSY-CA can address limitations of previous traditional meta-analyses such as adjusting for key confounders in all cohorts. Therefore, the results from PSY-CA are more reliable and interpretable.

Given the link of health behaviors such as smoking with both psychosocial factors (Strine et al., 2008; Verger et al., 2009) and cancer (Biswas et al., 2015; Chen et al., 2018; Kerr et al., 2017), there is a need to clarify whether the role of health behaviors is more than a confounding effect. Health behaviors, demographic and somatic factors that are well established cancer risk factors may interact with psychosocial factors to pose further risk. Furthermore, health behaviors could explain the link between psychosocial factors and cancer (i.e., health behaviors as mediators). Research into the role of health behaviors in the association between psychosocial factors and cancer is surprisingly lacking, and PSY-CA aims to provide insight into this area. As such, the results from the proposed study outlined in this article may reveal psychosocial factors that put individuals at risk for cancer, identify certain subgroups to target with preventive interventions, and support the use of health-behavior interventions to reduce the risk of cancer associated with psychosocial factors.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

ORCID

Lonneke A. van Tuijl https://orcid.org/0000-0001-9031-0886
Femke Lamers https://orcid.org/0000-0003-4344-5766

REFERENCES

Awadalla, P., Boileau, C., Payette, Y., Idaghdour, Y., Goulet, J.-P., Knoppers, B., Hamet, P., & Laberge, C. (2013). Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. *International Journal of Epidemiology, 42*(5), 1285–1299. https://doi.org/10.1093/ije/dys160

van den Belt-Dusebout, A. W., Spaan, M., Lambalk, C. B., Kortman, M., Laven, J. S. E., van Santbrink, E. J. P., van der Westerlaken, L. A. J., Cohlen, B. J., Braat, D. D. M., Smeenk, J. M. J., Land, J. A., Goddijn, M., van Golde, R. J. T., van Rumste, M. M., Schats, R., Józwiak, K., Hauptmann, M., Rookus, M. A., Burger, C. W., & van Leeuwen, F. E. (2016). Ovarian stimulation for in vitro fertilization and long-term risk of breast cancer. *Jama*, *316*(3), 300. https://doi.org/10.1001/jama.2016.9389

Benca, R. M., Okawa, M., Uchiyama, M., Ozaki, S., Nakajima, T., Shibui, K., & Obermeyer, W. H. (1997). Sleep and mood disorders. *Sleep Medicine Reviews*, *1*(1), 45–56. https://doi.org/10.1016/S1087-0792(97)90005-8

Beulens, J. W. J., Monninkhof, E. M., Monique Verschuren, W. M., van der Schouw, Y. T., Smit, J., Ocke, M. C., Jansen, E. H. J. M., van Dieren, S., Grobbee, D. E., Peeters, P. H. M., & Bueno-de-Mesquita, H. B. (2010). Cohort profile: The EPIC-NL study. *International Journal of Epidemiology*, *39*(5), 1170–1178. https://doi.org/10.1093/ije/dyp217

Biswas, A., Oh, P. I., Faulkner, G. E., Bajaj, R. R., Silver, M. A., Mitchell, M. S., & Alter, D. A. (2015). Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: A systematic review and meta-analysis. *Annals of Internal Medicine*, *162*(2), 123. https://doi.org/10.7326/M14-1651

Blaine, B. (2008). Does depression cause obesity? A meta-analysis of longitudinal studies of depression and weight control. *Journal of Health Psychology*, *13*(8), 1190–1197. https://doi.org/10.1177/1359105308095977

Boker, L. K., van Noord, P. A. H., van der Schouw, Y. T., Koot, N. V. C. M., Bas Bueno de Mesquita, H., Riboli, E., Grobbee, D. E., & Peeters, P. H. M. (2001). Prospect-EPIC Utrecht: Study design and characteristics of the cohort population. *European Journal of Epidemiology*, *17*(11), 1047–1053. https://doi.org/10.1023/A:1020009325797

Bono, C., Ried, L. D., Kimberlin, C., & Vogel, B. (2007). Missing data on the center for epidemiologic studies depression scale: A comparison of 4 imputation techniques. *Research in Social and Administrative Pharmacy*, *3*(1), 1–27. https://doi.org/10.1016/j.sapharm.2006.04.001. https://doi.org/10.1016/j.sapharm.2006.04.001

Borugian, M. J., Robson, P., Fortier, I., Parker, L., McLaughlin, J., Knoppers, B. M., Bedard, K., Gallagher, R. P., Sinclair, S., Ferretti, V., Whelan, H., Hoskin, D., & Potter, J. D. (2010). The Canadian partnership for tomorrow project: Building a Pan-Canadian research platform for disease prevention. *CMAJ: Canadian Medical Association Journal*, *182*(11), 1197–1201. https://doi.org/10.1503/cmaj.091540

Brookes, S. T., Whitely, E., Egger, M., Smith, G. D., Mulheran, P. A., & Peters, T. J. (2004). Subgroup analyses in randomized trials: Risks of subgroup-specific analyses; power and sample size for the interaction test. *Journal of Clinical Epidemiology*, *57*(3), 229–236. https://doi.org/10.1016/j.jclinepi.2003.08.009

Bulloch, A., Lavorato, D., Williams, J., & Patten, S. (2012). Alcohol consumption and major depression in the general population: The critical importance of dependence. *Depression and Anxiety*, *29*(12), 1058–1064. https://doi.org/10.1002/da.22001

Chen, Y., Tan, F., Wei, L., Li, X., Lyu, Z., Feng, X., Wen, Y., Guo, L., He, J., Dai, M., & Li, N.i. (2018). Sleep duration and the risk of cancer: A systematic review and meta-analysis including dose–response relationship. *BMC Cancer*, *18*(1), 1149. https://doi.org/10.1186/s12885-018-5025-y

Chida, Y., Hamer, M., Wardle, J., & Steptoe, A. (2008). Do stress-related psychosocial factors contribute to cancer incidence and survival? *Nature Clinical Practice Oncology*, *5*(8), 466. https://doi.org/10.1038/ncponc1134

Dalton, S. O., Boesen, E. H., Ross, L., Schapiro, I. R., & Johansen, C. (2002). Mind and cancer: Do psychological factors cause cancer? *European Journal of Cancer*, *38*(10), 1313–1323. https://doi.org/10.1016/S0959-8049(02)00099-0

Dummer, T. J. B., Awadalla, P., Boileau, C., Craig, C., Fortier, I., Goel, V., Hicks, J. M. T., Jacquemont, S., Knoppers, B. M., Le, N., McDonald, T., McLaughlin, J., Mes-Masson, A.-M., Nuyt, A.-M., Palmer, L. J., Parker, L., Purdue, M., Robson, P. J., Spinelli, J. J., … CPTP Regional Cohort Consortium.

(2018). The canadian partnership for tomorrow project: A pan-canadian platform for research on chronic disease prevention. *Canadian Medical Association Journal*, *190*(23), E710–E717. https://doi.org/10.1503/cmaj.170292

Ensor, J., Burke, D. L., Snell, K. I. E., Hemming, K., & Riley, R. D. (2018). Simulation-based power calculations for planning a two-stage individual participant data meta-analysis. *BMC Medical Research Methodology*, *18*(1), 41. https://doi.org/10.1186/s12874-018-0492-z

Fortier, I., Raina, P., Van den Heuvel, E. R., Griffith, L. E., Craig, C., Saliba, M., Doiron, D., Stolk, R. P., Knoppers, B. M., Ferretti, V., Granda, P., & Burton, P. (2016). Maelstrom research guidelines for rigorous retrospective data harmonization. *International Journal of Epidemiology*, *46*(1), 103–105. https://doi.org/10.1093/ije/dyw075

Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Smith, G. D., Henderson, J., Macleod, J., Molloy, L., Ness, A., Ring, S., Nelson, S. M., & Lawlor, D. A. (2013). Cohort profile: The avon longitudinal study of parents and children: ALSPAC mothers cohort. *International Journal of Epidemiology*, *42*(1), 97–110. https://doi.org/10.1093/ije/dys066

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *18*(3), 233–239. https://doi.org/10.1111/j.1467-9280.2007.01882.x

Grobbee, D. E., Hoes, A. W., Verheij, T. J. M., Schrijvers, A. J. P., van Ameijden, E. J. C., & Numans, M. E. (2005). The Utrecht health project: Optimization of routine healthcare data for research. *European Journal of Epidemiology*, *20*(3), 285–290. https://doi.org/10.1007/s10654-004-5689-2

Hofman, A., Brusselle, G. G. O., Murad, S. D., van Duijn, C. M., Franco, O. H., Goedegebure, A., Ikram, M. A., Klaver, C. C. W., Nijsten, T. E. C., Peeters, R. P., Stricker, B. H. C., Tiemeier, H. W., Uitterlinden, A. G., & Vernooij, M. W. (2015). The rotterdam study: 2016 objectives and design update. *European Journal of Epidemiology*, *30*(8), 661–708. https://doi.org/10.1007/s10654-015-0082-x

Hoogendijk, E. O., Deeg, D. J. H., Poppelaars, J., van der Horst, M., van Groenou, M. I. B., Comijs, H. C., Pasman, H. R. W., van Schoor, N. M., Suanet, B., Thomése, F., van Tilburg, T. G., Visser, M., & Huisman, M. (2016). The longitudinal aging study Amsterdam: Cohort update 2016 and major findings. *European Journal of Epidemiology*, *31*(9), 927–945. https://doi.org/10.1007/s10654-016-0192-0

Huisman, M., Poppelaars, J., van der Horst, M., Beekman, A. T., Brug, J., van Tilburg, T. G., & Deeg, D. J. (2011). Cohort profile: The longitudinal aging study Amsterdam. *International Journal of Epidemiology*, *40*(4), 868–876. https://doi.org/10.1093/ije/dyq219

Hurley, S., Goldberg, D., Bernstein, L., & Reynolds, P. (2015). Sleep duration and cancer risk in women. *Cancer Causes & Control*, *26*(7), 1037–1045.

Idahl, A., Hermansson, A., & Lalos, A. (2018). Social support and ovarian cancer incidence—A Swedish prospective population-based study. *Gynecologic Oncology*, *149*(2), 324–328. https://doi.org/10.1016/j.ygyno.2018.03.042

Internation Agency for Research on Cancer. (2019). *IARC monographs on the identification of carcinogenic hazards to humans and handbooks of cancer prevention.* https://monographs.iarc.fr/wp-content/uploads/2019/12/OrganSitePoster.PlusHandbooks.29112019.pdf.

Ji, X.u, Cummings, J. R., Marchak, J. G., Han, X., & Mertens, A. C. (2020). Mental health among nonelderly adult cancer survivors: A national estimate. *Cancer*, *126*(16), 3768–3776. https://doi.org/10.1002/cncr.32988

Jia, Y., Li, F., Liu, Y. F., Zhao, J. P., Leng, M. M., & Chen, L. (2017). Depression and cancer risk: A systematic review and meta-analysis. *Public Health*, *149*, 138–148. https://doi.org/10.1016/j.puhe.2017.04.026

Kerr, J., Anderson, C., & Lippman, S. M. (2017). Physical activity, sedentary behaviour, diet, and cancer: An update and emerging new evidence. *The Lancet Oncology*, *18*(8), e457–e471. https://doi.org/10.1016/S1470-2045(17)30411-4

Knekt, P., Hakama, M., Järvinen, R., Pukkala, E., & Heliövaara, M. (1998). Smoking and risk of colorectal cancer. *British Journal of Cancer*, *78*(1), 136–139. https://doi.org/10.1038/bjc.1998.455

Kroenke, C. H., Bennett, G. G., Fuchs, C., Giovannucci, E. D., Kawachi, I., Schernhammer, E., Holmes, M. D., & Kubzansky, L. D. (2005). Depressive symptoms and prospective incidence of colorectal cancer in women. *American Journal of Epidemiology*, *162*(9), 839–848. https://doi.org/10.1093/aje/kwi302

Krokstad, S., Langhammer, A., Hveem, K., Holmen, T. L., Midthjell, K., Stene, T. R., Bratberg, G., Heggland, J., & Holmen, J. (2013). Cohort profile: The HUNT study, Norway. *International Journal of Epidemiology*, *42*(4), 968–977. https://doi.org/10.1093/ije/dys095

Linkins, R. W., & Comstock, G. W. (1990). Depressed mood and development of cancer. *American Journal of Epidemiology*, *132*(5), 962–972. https://doi.org/10.1093/oxfordjournals.aje.a115739

Lutgendorf, S. K., Costanzo, E. S., & Siegel, S. D. (2007). Psychosocial influences in oncology: An expanded model of biobehavioral mechanisms. In R. Ader (Ed.), *Psychoneuroimmunology* (Vol. 2, pp. 869–895). Academic Press.

Marmot, M., & Brunner, E. (2005). Cohort profile: The whitehall II study. *International Journal of Epidemiology*, *34*(2), 251–256. https://doi.org/10.1093/ije/dyh372

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*(2), 376–390. https://doi.org/10.1037/0033-2909.114.2.376

McGee, R., Williams, S., & Elwood, M. (1994). Depression and the development of cancer: A meta-analysis. *Social Science & Medicine*, *38*(1), 187–192. https://doi.org/10.1016/0277-9536(94)90314-X

Mouw, T., Koster, A., Wright, M. E., Blank, M. M., Moore, S. C., Hollenbeck, A., & Schatzkin, A. (2008). Education and risk of cancer in a large cohort of men and women in the United States. *PLoS One*, *3*(11), e3639. https://doi.org/10.1371/journal.pone.0003639

O'Donovan, A., Hughes, B. M., Slavich, G. M., Lynch, L., Cronin, M.-T., O'Farrelly, C., & Malone, K. M. (2010). Clinical anxiety, cortisol and interleukin-6: Evidence for specificity in emotion–biology relationships. *Brain, Behavior, and Immunity*, *24*(7), 1074–1077. https://doi.org/10.1016/j.bbi.2010.03.003

Oerlemans, M. E., van den Akker, M., Schuurman, A. G., Kellen, E., & Buntinx, F. (2007). A meta-analysis on depression and subsequent cancer risk. *Clinical Practice and Epidemiology in Mental Health*, *3*, 29. https://doi.org/10.1186/1745-0179-3-29

Peled, R., Carmil, D., Siboni-Samocha, O., & Shoham-Vardi, I. (2008). Breast cancer, psychological distress and life events among young women. *BMC Cancer*, *8*(1), 245. https://doi.org/10.1186/1471-2407-8-245

Pelucchi, C., Tramacere, I., Boffetta, P., Negri, E., & Vecchia, C. L. (2011). Alcohol consumption and cancer risk. *Nutrition and Cancer*, *63*(7), 983–990. https://doi.org/10.1080/01635581.2011.596642

Penninx, B. W. J. H., Beekman, A. T. F., Johannes, H. S., Zitman, F. G., Nolen, W. A., Spinhoven, P., Cuijpers, P., de Jong, P. J., van Maruiijk, H. W. J., Assendelft, W. J. J., van der Meer, K., Verhaak, P., Wensing, M., de Graaf, R., Hoogendijk, W. J., Ormel, J., & van Dyck, R. (2008). The Netherlands Study of Depression and Anxiety (NESDA): Rationale, objectives and methods. *International Journal of Methods in Psychiatric Research*, *17*(3), 121–140. https://doi.org/10.1002/mpr.256

Randi, G., Altieri, A., Gallus, S., Chatenoud, L., Montella, M., Franceschi, S., Negri, E., Talamini, R., & Vecchia, C. L.a. (2004). Marital status and cancer risk in Italy. *Preventive Medicine*, *38*(5), 523–528. https://doi.org/10.1016/j.ypmed.2003.12.004

Riboli, E. (1992). Nutrition and cancer: Background and rationale of the european prospective investigation into cancer and nutrition (EPIC). *Annals of Oncology*, *3*(10), 783–791. https://doi.org/10.1093/oxfordjournals.annonc.a058097

Riboli, E., Hunt, K. J., Slimani, N., Ferrari, P., Norat, T., Fahey, M., Charrondière, U. R., Hémon, B., Casagrande, C., Vignat, J., Overvad, K., Tjønneland, A., Clavel-Chapelon, F., Thiébaut, A., Wahrendorf, J., Boeing, H., Trichopoulos, D., Trichopoulou, A., Vineis, P., … Saracci, R. (2002). European prospective investigation into cancer and nutrition (EPIC): Study populations and data collection. *Public Health Nutrition*, *5*(6b), 1113–1124. https://doi.org/10.1079/PHN2002394

Roshanaei-Moghaddam, B., Katon, W. J., & Russo, J. (2009). The longitudinal effects of depression on physical activity. *General Hospital Psychiatry*, *31*(4), 306–315. https://doi.org/10.1016/j.genhosppsych.2009.04.002

Schapiro, I. R., Ross-Petersen, L., Sælan, H., Garde, K., Olsen, J. H., & Johansen, C. (2001). Extroversion and neuroticism and the associated risk of cancer: A Danish cohort study. *American Journal of Epidemiology*, *153*(8), 757–763. https://doi.org/10.1093/aje/153.8.757

Scholtens, S., Smidt, N., Swertz, M. A., Bakker, S. J. L., Dotinga, A., Vonk, J. M., Dijk, F., van Zon, S. K. R., Wijmenga, C., Wolffenbuttel, B. H. R., & Stolk, R. P. (2015). Cohort profile: LifeLines, a three-generation cohort study and biobank. *International Journal of Epidemiology*, *44*(4), 1172–1180. https://doi.org/10.1093/ije/dyu229

Shrive, F. M., Stuart, H., Quan, H., & Ghali, W. A. (2006). Dealing with missing data in a multi-question depression scale: A comparison of imputation methods. *BMC Medical Research Methodology*, *6*(1), 57. https://doi.org/10.1186/1471-2288-6-57

Simons, P. C. G., Algra, A., van de Laak, M. F., Grobbee, D. E., van der Graaf, Y., & SMART Study Group. (1999). Second manifestations of ARTerial disease (SMART) study: Rationale and design. *European Journal of Epidemiology*, *15*(9), 773–781. https://doi.org/10.1023/A:1007621514757

Slottje, P., Yzermans, C. J., Korevaar, J. C., Hooiveld, M., & Vermeulen, R. C. H. (2014). The population-based occupational and environmental health prospective cohort study (AMIGO) in the Netherlands. *BMJ Open*, *4*(11), e005858. https://doi.org/10.1136/bmjopen-2014-005858

Snijder, M. B., Galenkamp, H., Prins, M., Derks, E. M., Peters, R. J. G., Zwinderman, A. H., & Stronks, K. (2017). Cohort profile: The healthy life in an urban setting (HELIUS) study in Amsterdam, The Netherlands. *BMJ Open*, *7*(12), e017873. https://doi.org/10.1136/bmjopen-2017-017873

Spaan, M., van den Belt-Dusebout, A. W., Burger, C. W., van Leeuwen, F. E., Schats, R., Lambalk, C. B., Kortman, M., Laven, J. S. E., Jansen, C. A. M., van der Westerlaken, L. A. J., Cohlen, B. J., Braat, D. D. M., Smeenk, J. M. J., Land, J. A., van der Veen, F., Evers, J. L. H., & van Rumste, M. M. E. (2016). Risk of colorectal cancer after ovarian stimulation for in vitro fertilization. *Clinical Gastroenterology and Hepatology*, *14*(5), 729–737.e5. https://doi.org/10.1016/j.cgh.2015.12.018

Steptoe, A., Breeze, E., Banks, J., & Nazroo, J. (2013). Cohort profile: The English longitudinal study of ageing. *International Journal of Epidemiology*, *42*(6), 1640–1648. https://doi.org/10.1093/ije/dys168

Steuber, T. L., & Danner, F. (2006). Adolescent smoking and depression: Which comes first? *Addictive Behaviors*, *31*(1), 133–136. https://doi.org/10.1016/j.addbeh.2005.04.010

Stewart, L. A., & Parmar, M. K. B. (1993). Meta-analysis of the literature or of individual patient data: Is there a difference?' *The Lancet*, *341*(8842), 418–422. https://doi.org/10.1016/0140-6736(93)93004-K

Strine, T. W., Mokdad, A. H., Dube, S. R., Balluz, L. S., Gonzalez, O., Berry, J. T., Manderscheid, R., & Kroenke, K. (2008). The association of depression and anxiety with obesity and unhealthy behaviors among community-dwelling US adults. *General Hospital Psychiatry*, *30*(2), 127–137. https://doi.org/10.1016/j.genhosppsych.2007.12.008

Stronks, K., Snijder, M. B., Peters, R. J. G., Prins, M., Schene, A. H., & Zwinderman, A. H. (2013). Unravelling the impact of ethnicity on health in europe: The HELIUS study. *BMC Public Health*, *13*(1), 402. https://doi.org/10.1186/1471-2458-13-402

Sweeney, E., Cui, Y., DeClercq, V., Devichand, P., Forbes, C., Grandy, S., Hicks, J. M. T., Keats, M., Parker, L., Thompson, D., Volodarsky, M., Yu, Z. M., & Dummer, T. J. B. (2017). Cohort profile: The atlantic partnership for tomorrow's health (Atlantic PATH) study. *International Journal of Epidemiology*, *46*(6), 1762–1763i. https://doi.org/10.1093/ije/dyx124

Thakkar, J. P., Villano, J. L., & McCarthy, B. J. (2014). Age-specific cancer incidence rates increase through the oldest age groups. *The American Journal of the Medical Sciences*, *348*(1), 65–70. https://doi.org/10.1097/MAJ.0000000000000281

Tierney, J. F., Stewart, L. A., & Clarke, M. (2020). Chapter 26: Individual participant data. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021)*. Cochrane. Available from www.training.cochrane.org/handbook.

Trichopoulos, D., MacMahon, B., & Cole, P. (1972). Menopause and breast cancer risk. *JNCI: Journal of the National Cancer Institute*, *48*(3), 605–613. https://doi.org/10.1093/jnci/48.3.605

Vecchia, C. L., Brinton, L. A., & McTiernan, A. (2001). Menopause, hormone replacement therapy and cancer. *Maturitas*, *39*(2), 97–115. https://doi.org/10.1016/S0378-5122(01)00213-4

Verger, P., Lions, C., & Ventelou, B. (2009). Is depression associated with health risk-related behaviour clusters in adults? *European Journal of Public Health*, *19*(6), 618–624. https://doi.org/10.1093/eurpub/ckp057

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36): I. conceptual framework and item selection. *Medical Care*, *30*(6), 473–483. https://doi.org/10.1097/00005650-199206000-00002

White, A., Ironmonger, L., Steele, R. J. C., Ormiston-Smith, N., Crawford, C., & Seims, A. (2018). A review of sex-related differences in colorectal cancer incidence, screening uptake, routes to diagnosis, cancer stage and survival in the UK. *BMC Cancer*, *18*(1), 906. https://doi.org/10.1186/s12885-018-4786-7

APPENDIX 1

Specific hypotheses by research questions

Research question 1: Do psychosocial factors (depression, anxiety, perceived lack of social support, relationship status, recent loss events, neuroticism, and general distress) increase the incidence of cancer? Specifically, we hypothesize that depression, anxiety, recent loss events, and perceived lack of social support all individually increase the incidence of any cancer, breast cancer, lung cancer, prostate cancer, colorectal cancer, smoking-related cancers, and alcohol-related cancers. We limit our hypotheses to depression, anxiety, recent loss events and perceived low social support given the rather clear distinction between these concepts (e.g., while neuroticism and general distress are relatively broad constructs incorporating symptoms of both depression and anxiety), and the focus on these factors in prior research (e.g., relatively little research has looked at relationship status and cancer incidence). Therefore, analyses relating to neuroticism, general distress and relationship status are considered explorative.

Research question 2: Do these psychosocial factors interact with health behaviors (smoking, alcohol use, weight, physical activity, sedentary behavior, sleep duration, and sleep quality) or demographic and clinical factors (age, sex, education, hormone replacement therapy, and menopausal status) on the risk of cancer incidence? Specifically, we hypothesize that the risk of cancer in people with psychosocial stress (i.e., elevated depression symptom level or diagnosis, elevated anxiety symptom level or diagnosis, a recent loss event, or perceived lack of social support) and unhealthy behavior (smoking, alcohol use, overweight, and low physical activity) is greater than the sum of the individual effects of the psychosocial factor and unhealthy behavior on cancer incidence. We limit our hypotheses to these health-related behaviors given the consistent evidence of their association with cancer.

Research question 3: Are the relationships between these psychosocial factors and incidence of cancer mediated by health-related factors (smoking, alcohol use, weight, physical activity, sedentary behavior, sleep duration, and sleep quality)? Again limiting the hypotheses to depression, anxiety, recent loss events, and perceived low social support, we hypothesize that (a) smoking, alcohol use, physical inactivity, and high body mass index (BMI) partially mediate the association between the psychosocial factors and cancer of any kind, breast cancer, and colorectal cancer; (b) smoking partially mediates the association between the psychosocial factors and smoking-related cancers; (c) alcohol use partially mediates the association between the psychosocial factors and alcohol-related cancers; (d) smoking and physical inactivity partially mediate the association between the psychosocial factors and lung cancer; and (e) physical inactivity partially mediates the association between the psychosocial factors and prostate cancer.