The impact of health behaviors on incident and recurrent cancers: a population based analysis
Cortes Ibañez, Francisco Omar

DOI: 10.33612/diss.231240920

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

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

Publication date: 2022

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Chapter 1                                      General introduction

Chapter 1

Cancer epidemiology

The epidemiological transition from communicable to non-communicable diseases has challenged the world’s population well-being in the last three decades (1–3). In 2016, non-communicable diseases (cancer, cardiovascular, diabetes and chronic respiratory problems) accounted for 71% (41 million) of the total mortality in the world (57 million) (4). According to the report of World Health Organization (WHO), malignant tumors are the leading cause of global death, in 2020 only about 10 million people died due to oncological diseases (5). Along with this, cancer incidence has steadily increased worldwide; rising from 8.1 million in 1999 to 19.3 million in 2020 (5–9). In addition, the number of people living globally with a cancer diagnosis (cancer survivors) was estimated to be around 44 million in 2018 (10). The most commonly diagnosed cancers in 2020 were breast (2.26 million), lung (2.20 million), prostate (1.41 million), skin (1.19 million), colon (1.14 million) and stomach cancer (1.08 million). Together, they represent around 50% of all the cancer diagnoses in the world, and one third of total mortality, in which lung cancer was leading with 1.79 million deaths for the same year (5).

What causes cancer?

Aging is the most important factor contributing to the occurrence of cancer. Adults aged 65 years or older accounted for 50% of the cancer diagnoses worldwide in 2020, and this is expected to increase to 60% from the total diagnoses by 2034 (5,11). Even though several socioeconomic, lifestyle, genetic and environmental factors are shown to be associated with the development of malignancies (1,8,12), this thesis will focus on evaluating health behaviours of people and how these are related to the risk of development and recurrence of cancers. The motivation for this is that an estimated 40% of the overall burden of new cancer cases could be reduced through lifestyle modification (13).

Risk factors for cancer

Evidence about the role of different health behaviours in cancer occurrence has been growing in recent years. However, many studies are limited to an individual assessment of potential risk factors and do not evaluate the combined effect of the variety of health behaviours on the occurrence of cancer. Health behaviours can interact differently across...
Cancer epidemiology

The epidemiological transition from communicable to non-communicable diseases has challenged the world’s population well-being in the last three decades (1–3). In 2016, non-communicable diseases (cancer, cardiovascular, diabetes and chronic respiratory problems) accounted for 71% (41 million) of the total mortality in the world (57 million) (4). According to the report of World Health Organization (WHO), malignant tumors are the leading cause of global death, in 2020 only about 10 million people died due to oncological diseases (5). Along with this, cancer incidence has steadily increased worldwide; rising from 8.1 million in 1999 to 19.3 million in 2020 (5–9). In addition, the number of people living globally with a cancer diagnosis (cancer survivors) was estimated to be around 44 million in 2018 (10). The most commonly diagnosed cancers in 2020 were breast (2.26 million), lung (2.20 million), prostate (1.41 million), skin (1.19 million), colon (1.14 million) and stomach cancer (1.08 million). Together, they represent around 50% of all the cancer diagnoses in the world, and one third of total mortality, in which lung cancer was leading with 1.79 million deaths for the same year (5).

What causes cancer?

Aging is the most important factor contributing to the occurrence of cancer. Adults aged 65 years or older accounted for 50% of the cancer diagnoses worldwide in 2020, and this is expected to increase to 60% from the total diagnoses by 2034 (5,11). Even though several socioeconomic, lifestyle, genetic and environmental factors are shown to be associated with the development of malignancies (1,8,12), this thesis will focus on evaluating health behaviours of people and how these are related to the risk of development and recurrence of cancers. The motivation for this is that an estimated 40% of the overall burden of new cancer cases could be reduced through lifestyle modification (13).

Risk factors for cancer

Evidence about the role of different health behaviours in cancer occurrence has been growing in recent years. However, many studies are limited to an individual assessment of potential risk factors and do not evaluate the combined effect of the variety of health behaviours on the occurrence of cancer. Health behaviours can interact differently across
populations, thereby contributing to accelerating cancer incidence and mortality (1). Smoking and alcohol consumption are of particular interest because these are known risk factors of several cancer types. Smoking is known to be strongly related to lung cancer (14). Smoking and alcohol consumption are also linked to other cancers, where the most common ones (among others) are colorectal, liver, stomach, oesophagus and head & neck cancers, see Figure 1 (15–17).

Fig 1. Cancers associated to tobacco smoking and alcohol consumption, as shown in Anand et al. (16).

Other factors that play an important role in an increased cancer risk are: unhealthy diet, physical inactivity, body mass index (BMI) and sedentary behaviour (17). An unhealthy diet is shown to be associated with an increased risk of colon cancer (18). Especically red meat consumption is identified as a risk factor for developing gastrointestinal, breast, pancreatic and bladder cancers (19). Moreover, higher levels of physical activity are associated with lower risk for 13 out of 26 types of cancer, where breast and colo-rectal cancers are the most common ones (20). Furthermore, a high BMI is associated with six types of cancer including pancreatic, endometrial and rectal cancer (21). Recent evidence suggests that sedentary behaviour should be evaluated separately from the overall physical activity; in those studies, sedentary behaviour is strongly related to colorectal and endometrial cancer (22,23).
Chapter 1

**What can be done about it?**

It has been estimated that 40% of the overall burden of new cancer cases could be reduced by health behaviours modification (i.e., tobacco smoking reduction or change on dietary patterns) (13), or a change in the exposure to environmental factors (24). But evidence for this is still inconclusive and should be investigated in other populations (13,24). In addition, effective treatments, and the early detection by the implementation of screening programs got a lot of attention because they are intended to accomplish longer survival after a cancer diagnosis (10). However, there is still a big gap between low-middle income countries versus high income countries on the access to new cancer treatments (25) and the implementation of such screening programs with a focus on the early detection of cancer (26,27).

There is evidence supporting that oncological diseases continue to be a major public health problem, that persists after recovery from a primary cancer. Recent population-based studies evaluated differences in health behaviours of cancer survivors as compared to people without any history of cancer (28–36). Those studies were mainly focused on the evaluation of a single health behaviour in cancer survivors, such as smoking or physical activity. A few studies looked at more than one health behaviour in the same population, typically adjusting for age, sex and various sets of other potential confounders. In this thesis, we focused on evaluating the impact of multiple health behaviours on the risk of incident cancer and cancer recurrence.

**Different methods, different outcomes?**

In cancer epidemiology research, the most common methods for evaluating an association between a risk factor and an outcome is linear regression, or its variations (37,38). Such analyses often depend on the distribution of the data, hence some assumptions need to be fulfilled (i.e. linearity and normal distribution) and cannot handle complex or non-linear associations between health behaviours. It is hypothesized that machine learning algorithms (MLAs) will be able to identify potentially relevant but undetected associations between health behaviours and cancers in aging-related research (39). These algorithms are designed to handle non-linear relationships between variables, as well as large number of variables that may increase the classification/prediction performance compared to traditional statistical
approaches (40–45). Therefore, in this thesis we used and compared the performance of the traditional linear approaches with the non-linear MLAs for the evaluation of the associations between the health beahaviours and the risk of cancer occurrence and recurrence.

**Lifelines and PALGA**

The studies within this thesis are based on data from the Lifelines population-based cohort. LifeLines is a unique, large and prospective databank designed to study the etiology of health and healthy ageing that includes 10% of the population from the north of The Netherlands (46). Lifelines comprehends a three-generation design; starting in 2006, the 167,729 participants included in Lifelines provided extensive information about their health, quality of life, lifestyle, socioeconomic status and, among others, their cancer history. Part of the information collected from Lifelines is obtained through validated questionnaires, covering several health domains from the participants, and from self-reported data. Also the information on the occurrence of cancer is based on self-reporting. From previous studies it is known that self-reported information for health domains tends to be biased and needs to be validated (47). Such validations may allow researchers to use and to better understand this self-reported data. Literature with regard to the validation of self-reported cancer data is scarce and limited to a few population based studies (48–51). To validate the self-reported cancer data from Lifelines, a gold standard source on cancer data: a nationwide pathology registry (PALGA, Pathological Anatomical National Automated Archive) was used. This registry records cancer diagnoses within the Netherlands and has full coverage (52). For this study, data from PALGA were linked to data from Lifelines.
Chapter 1

THESIS OUTLINE

Aim of the thesis

The primary aim of this thesis was to investigate the role of health behaviours in both incident cancer diagnosis and cancer survivors. Moreover, this thesis also assessed if previously reported associations between health behaviours and cancer incidence using traditional approaches, can be outperformed by machine learning algorithms.

Part 1. Validation of self-reported cancer

In chapter 2 the validity of self-reported cancer diagnoses as reported in the baseline assessment of the Lifelines cohort were compared to the cancer diagnoses in the PALGA registry which was considered as a gold standard.

Part 2. Lifestyle as a predictor of incident cancer

In chapter 3 the performance of the traditional linear approaches and non-linear machine learning algorithms (random forest, support vector machines and logistic regression) in the assessment of associations between baseline health behaviours and incident cancer outcome was compared in a large population based cohort.

In chapter 4, the influence of specific dietary patterns established in four different dietary indices (Dutch Dietary Guidelines (DDG) index (53), the Lifelines Diet Score (LLDS) (54), the American Cancer Society (ACS) index (55), the World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) index (56)) on long term gastrointestinal cancer risk was assessed.

Part 3. Lifestyle in cancer survivors

In chapter 5 it was evaluated if the lifestyle behaviours (alcohol drinking, smoking, body mass index, diet, physical activity and sedentary behaviour) in cancer survivors differ from those of the general population by means of logistic regression using complete case analysis.

In chapter 6, similar to in chapter 4 the performance of traditional linear methods (logistic regression) when classifying non-cancer history individuals opposed to those with a
history of cancer based on their health behaviours was compared to machine learning algorithms (random forest, support vector machines (SVM) and gradient boosting machines (GBM)).

Finally, in chapter 7, diet quality differences between gastrointestinal cancer survivors and the general population were evaluated by using matched controls, thereby adjusting for age, sex and educational level. Diet quality was assessed by the Lifelines diet score and its food components.
Finally, in chapter 7, diet quality differences between gastrointestinal cancer survivors and the general population were evaluated by using matched controls, thereby adjusting for age, sex and educational level. Diet quality was assessed by the Lifelines diet score and its food components.

References


