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):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 8                                        General Discussion

As the leading cause of morbidity and mortality in the world, cancer is among the biggest threats to population wellbeing (1). Lifestyle modification is estimated to potentially reduce the burden of cancer by around 40%; this encompasses reducing or quitting smoking (2) and alcohol drinking (3), to maintain a normal body mass index (BMI) (4), to increase the quality of diet (5) and to increase moderate to intense physical activity per week (6,7), in addition to reducing sedentary behaviour (8,9). However, the evidence is still inconclusive that all the health behaviours mentioned above are indeed related to a cancer incidence. For factors like diet (10–12), physical activity (6), sedentary behaviour (8,9) and BMI (4) the recommendation is that further evaluation is necessary, particularly in population-based cohorts, because some geographical or cultural differences might have not been explored yet. In addition, there is evidence suggesting that risk health behaviours persists even after a cancer diagnosis, and that cancer survivors still need to improve certain aspects of their lifestyle to avoid cancer recurrences, therefore further studies are needed to substantiate the previous findings (13,14). The aim of this thesis therefore was to investigate the role of health behaviours in both incident cancer diagnosis and cancer survivors. Moreover, it was explored if the associations reported for the health behaviours in previous literature relying on traditional approaches, could also be found by machine learning algorithms. The discussion chapter will first describe the main findings per chapter of this thesis, followed by a summary of the results by each health behaviour compared to previous evidence. In addition, the methodological, research, and clinical implications will be discussed. The final part of this chapter will provide thoughts about future perspectives of this work and draw general conclusions.

In part 1 of this thesis, chapter 2 focuses on assessing the validity of self-reported cancer diagnoses as provided by participants of Lifelines, which were compared to the cancer diagnoses as recorded by the Dutch pathology registry, PALGA Foundation (Pathological Anatomical National Automated Archive), which was considered as a gold standard. The study found that the self-reported cancer diagnoses by the participants in the Lifelines population-based cohort have high positive predictive value (97.45% [95%CI: 97.05–97.81], and 97.33% [95%CI: 96.72–97.82] after the exclusion of skin and cervical cancer) and mode rate sensitivity (64.7% [95%CI: 63.7–65.6]). The sensitivity increased when cervical and skin cancer were excluded (70.2% [95%CI: 68.8–71.5]) due to the observation that these two cancers were underreported. The major strengths of that study is that it is the first study in the Netherlands and among the largest population-based studies in the world validating self-reported cancer diagnoses (15–18). In addition, the important advantage of the present study is that a nationwide pathology registry,
General aim of the thesis

As the leading cause of morbidity and mortality in the world, cancer is among the biggest threats to population wellbeing (1). Lifestyle modification is estimated to potentially reduce the burden of cancer by around 40%; this encompasses reducing or quitting smoking (2) and alcohol drinking (3), to maintain a normal body mass index (BMI) (4), to increase the quality of diet (5) and to increase moderate to intense physical activity per week (6,7), in addition to reducing sedentary behaviour (8,9). However, the evidence is still inconclusive that all the health behaviours mentioned above are indeed related to a cancer incidence. For factors like diet (10–12), physical activity (6), sedentary behaviour (8,9) and BMI (4) the recommendation is that further evaluation is necessary, particularly in population-based cohorts, because some geographical or cultural differences might have not been explored yet. In addition, there is evidence suggesting that risk health behaviours persists even after a cancer diagnosis, and that cancer survivors still need to improve certain aspects of their lifestyle to avoid cancer recurrences, therefore further studies are needed to substantiate the previous findings (13,14). The aim of this thesis therefore was to investigate the role of health behaviours in both incident cancer diagnosis and cancer survivors. Moreover, it was explored if the associations reported for the health behaviours in previous literature relying on traditional approaches, could also be found by machine learning algorithms. The discussion chapter will first describe the main findings per chapter of this thesis, followed by a summary of the results by each health behaviour compared to previous evidence. In addition, the methodological, research, and clinical implications will be discussed. The final part of this chapter will provide thoughts about future perspectives of this work and draw general conclusions.

In part 1 of this thesis, chapter 2 focuses on assessing the validity of self-reported cancer diagnoses as provided by participants of Lifelines, which were compared to the cancer diagnoses as recorded by the Dutch pathology registry, PALGA Foundation (Pathological Anatomical National Automated Archive), which was considered as a gold standard. The study found that the self-reported cancer diagnoses by the participants in the Lifelines population-based cohort have high positive predictive value (97.45% [95%CI: 97.05-97.81], and 97.33% [95%CI: 96.72-97.82] after the exclusion of skin and cervical cancer) and moderate sensitivity (64.7% [95%CI:63.7-65.6]). The sensitivity increased when cervical and skin cancer were excluded (70.2% [95%CI:68.8-71.5]) due to the observation that these two cancers were underreported. The major strengths of that study is that it is the first study in the Netherlands and among the largest population-based studies in the world validating self-reported cancer diagnoses (15–18). In addition, the important advantage of the present study is that a nationwide pathology registry,
which has a full coverage of the pathology reports within the Netherlands, was used as a gold standard (19). The results of this study are consistent to those other studies from large population based-cohorts in Korea and the United States which showed similar sensitivity of their cancer diagnoses (72% [95%CI:70.7-73.2] and 79% respectively) (16,17). These results suggest that self-reported cancer diagnoses from the Lifelines population-based cohort can be used for research, with the exceptions for skin and cervical cancer because those were the main source of underreported cases.

The role of health behaviours on incident cancers was explored in part 2 of this thesis. Two different statistical approaches were applied to assess the relationships between health behaviours and incident cancers. (I) In chapter 3, machine learning algorithms were applied to explore potential non-linear associations between health behaviours and incident cancer diagnoses, and (II) in chapter 4 traditional survival analysis was used to assess the risk specifically for gastrointestinal cancers.

In chapter 3, the performance of two machine learning algorithms was compared to traditional linear methods in predicting incident cancer diagnoses (overall, breast, prostate and gastrointestinal). The machine learning algorithms were applied to evaluate if those could identify different patterns among lifestyle, socioeconomic factors and routine laboratory measurements as compared to traditional linear algorithms when predicting incident cancers. The results showed that linear and machine learning algorithms had comparable performance in cancers prediction for this population-based cohort, demonstrating an overall area under the receiver operator curve (AUC) of <0.75. The highest AUC was reported for prostate cancer (AUC>0.81). In general, the models were strongly dependent on age of participants, that explained most of the predictive performance. When excluding age from the models, only the specific model for gastrointestinal cancers showed low-to-moderate predictive performance (AUC=0.60) with smoking as the main contributor to the model.

In chapter 4 the role of different diet indices to gastrointestinal cancer risk were explored. The following four dietary indices were evaluated: the Dutch Dietary Guidelines (DDG) index (20), the Lifelines Diet Score (LLDS) (21), the American Cancer Society (ACS) index (22), the World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) index (23). The high diet quality assessed by the ACS index was associated with a reduced risk of colorectal cancer (HR 0.62; 95% CI 0.46–0.84). Since ACS index includes lifestyle components (physical activity and BMI) in its calculations, a sensitivity analysis was conducted that excluded those lifestyle components, thereby focussing on dietary components. The dietary components alone from the ACS, still showed to be significantly associated with a reduced risk of colorectal cancer (HR 0.68; 95% CI 0.49–0.93).
In part 3 of this thesis, the health behaviours of cancer survivors were compared to those participants with no-history of cancer from the Lifelines cohort. Chapter 5 evaluated the differences in health behaviours between cancer survivors and the general population by using logistic regression models, adjusted and stratified by sex and age groups. It is important to clarify here that, in the analyses of chapter 5 the reference category for the different lifestyle factors in each logistic regression was always the category indicating healthy behaviour. As such, when compared with the cancer-free group, a resulting odds ratio (OR) above 1 indicated that CSs were at increased odds for the unhealthy behaviour, while ORs below 1 indicated that CSs were at increased odds for the healthy behaviour. The following health behaviours were considered: smoking, alcohol intake, dietary quality, BMI, physical activity and sedentary behaviour. Overall, cancer survivors had a lower BMI (0.93 [95%CI:0.87-0.98]) and higher levels of physical activity (0.92 [95%CI:0.87-0.97]). To further evaluate this, stratified analyses by sex and age groups (under 55 years old and 55 years or older) were performed. Female cancer survivors were more likely to be current smokers compared to those females with no-history of cancer (1.14 [95%CI:1.04-1.26]). Male cancer survivors were more likely to have a normal BMI (0.87 [95%CI:0.78-0.97]) and be more physically active (0.86 [95%CI:0.78-0.95]) compared to males without a cancer diagnosis. Cancer survivors under the age of 55 were more likely to be current smokers (1.13 [95%CI:1.02-1.25]) and have a healthier diet (0.89 [95%CI:0.81-0.97]) compared to those in the same age group with no-cancer history. Cancer survivors in the Lifelines cohort still engage in risk health behaviours, especially female cancer survivors and those under the age of 55.

Chapter 6 evaluated if machine learning algorithms could better identify differences in health behaviours between cancer survivors and the non-history of cancer participants within the Lifelines cohort was compared to linear algorithms. The classification performance of logistic regression as a linear approach was compared to those of random forest, support vector machines (SVM), gradient boosting machines (GBM) as non-linear algorithms. Linear (logistic regression) and non-linear algorithms (random forest, support vector machines and gradient boosting machines) showed similar classification performance (logistic regression AUC=0.75, random forest AUC=0.75, support vector machines AUC=0.76 and gradient boosting machines AUC=0.74) when differentiating cancer survivors from those without a cancer diagnosis. In all the models age was the strongest classifier. When excluding age from the analyses all the classification performances decreased to AUC ≤ 0.66. These results suggest that linear and non-linear approaches did not perform differently in identifying key health behaviours to classify cancer survivors compared to non-history of cancer participants.
Finally, chapter 7 aimed to evaluate potential differences in adherence to dietary guidelines between gastrointestinal cancer survivors and healthy controls. Participants from the Lifelines cohort ≥40 years with a self-reported diagnosis of bowel, oesophageal or gastric cancer (n=307) were matched by age and sex to controls without a history of cancer (n=3070). Dietary intake was quantified using the Lifelines diet score which distinguishes between food groups according to their known health effect, positive (i.e., fruits and vegetables) or negative (i.e., red and processed meat) (21). The adherence to diet quality recommendations from Dutch dietary guidelines (20) was compared between cancer survivors and those without a cancer diagnosis. The recommended intake of fruits and vegetables per day (200 grams) was below 10% and 40% respectively. This demonstrated poor diet quality for both cancer survivors and those with no history of cancer. From the food groups with a known positive effect on health, only fish and whole grain products were consumed by cancer survivors and controls in amounts above the recommendations of the Dutch dietary guidelines, by 40% and 20% respectively. Hence, diet quality in gastrointestinal cancer survivors did not differ from that of those without cancer, meaning that both groups had poor adherence to the recommendations of Dutch dietary guidelines; something that could be improved.

Health behaviours and cancer

As already has been mentioned, there is consistent evidence for the relation between modifiable health behaviour such as smoking and alcohol intake on the one hand and the occurrence of cancer on the other hand. However, there is a lack of such evidence for the relation between modifiable health behaviours such as diet, physical activity, sedentary behaviour and BMI in relation to the incidence of cancer. The evidence about the role of these modifiable health behaviours in cancer occurrence will be shortly reviewed and compared to the findings of this thesis.

Smoking

Tobacco smoking has been described as a risk factor for several health complications, especially for lung diseases such as chronic obstructive pulmonary disease (COPD) and lung cancer, but also for gastrointestinal cancer (24). In Lifelines, tobacco smoking was measured for all the participants, before or after developing a malignancy (cancer survivors). In chapter 5, it was reported that half of the participants were either current or former smokers. Noteworthy, female cancer survivors were more likely to be current smokers compared to those with no history of cancer. A possible explanation for this might be their relatively long survival: on average nine years for all the cancer survivors. For these female cancer survivors, as their survival extends, they may eventually revert their smoking behaviour
(14). Those findings are consistent with the existing literature in which a cancer diagnosis is considered as a “teaching moment” (25) and cancer survivors convert to a better lifestyle when recently diagnosed, but eventually as the survival extends, they return to the lifestyle as prior to the diagnosis (14). In chapter 3, smoking was shown to be an important contributor for incident gastrointestinal cancers. Although the model for prediction of gastrointestinal cancers proposed in chapter 3 still has to be externally validated, the results are consistent with current literature about the relation between smoking and gastrointestinal cancer development (24).

Alcohol intake

The consumption of alcohol has been described as a risk factor for specific types of cancer, mainly in the gastrointestinal tract (esophageal, colorectal, but also in laryngeal, pharyngeal and breast cancer) (3). Around 70% of the participants in the Lifelines cohort reported at least one gram of alcohol intake per day. In chapter 5 no relevant differences were found between cancer survivors and non-history of cancer participants regarding alcohol consumption. Only the unadjusted analysis showed a tendency of cancer survivors to consume less alcohol compared to non-history of cancer participants. In the stratified analysis male cancer survivors also tended to consume less alcohol compared to non-history of cancer participants, but this result was non-significant. These results suggest that there is a tendency towards reduced alcohol drinking among cancer survivors, but not strong enough to be statistically significant. This is in line with a recent meta-analysis which showed lower alcohol intake in cancer survivors compared to the general population (14). Other studies in Denmark (26) and the United States (27) did not report differences between cancer survivors and general population with regard to alcohol consumption. Chapter 3 included alcohol consumption in the linear and non-linear analyses to evaluate its role in the prediction of the incidence of cancer overall, as well as separately for breast, prostate and gastrointestinal cancers. It was observed that in the proposed models, alcohol intake was not a relevant contributor to the classification performance. Hence, there was no substantial difference in alcohol consumption between participants who had incident cancers and those without a cancer diagnosis. This could be due to the relatively short follow-up time that was on average seven years and the level of exposure to alcohol was a single measurement. It has been shown that alcohol consumption should be evaluated in the long term to assess more accurately the association to cancer (28,29). In addition, the questionnaire used by Lifelines is susceptible to social and recall bias which might have hindered the amount of alcohol reported by the participants.

Diet
Diet quality was comprehensively assessed in the Lifelines cohort. The development of the Lifelines diet score allowed an evaluation of food intake in accordance with the Dutch dietary guidelines, not only by fruits and vegetables, but by a wider variety of food components (21). In chapter 5, 6 and 7 of this thesis, the diet quality of cancer survivors was compared to that of those without a cancer diagnosis. In chapter 5, overall logistic regression analyses, adjusted by age and sex did not show any substantial differences in diet quality between the two groups. Stratified analysis showed a better diet quality only in younger cancer survivors based on the Lifelines diet score. In chapter 6, the use of non/linear approaches also did not identify substantial differences in diet quality between cancer survivors and participants with no-history of cancer. In chapter 7, diet quality was compared solely between gastrointestinal cancer survivors and controls with no-history of cancer. It was observed that the diet quality intake was comparable between cancer survivors and those without cancer. Both gastrointestinal cancer survivors and healthy controls did not adhere to dietary recommendations for daily intake according to Dutch guidelines. The amount of food with a known positive health effect such as fruits, vegetables, legumes and nuts were consumed below the minimum amount suggested by the Dutch guidelines in both groups. But the daily intake goals of fish and whole grain products, which are considered healthy foods, were met.

The results described in chapters 5, 6 and 7 of this thesis, evaluating diet differences between cancer survivors and those without a cancer diagnosis are in line with other studies that were unable to find differences between cancer survivors and those without cancer (30,31). The findings of this thesis and in current literature show that diet quality is poor, not only in cancer survivors (13,32,33), but also in participants without a cancer diagnosis. In contrast, one previous study reported a higher consumption of fruits and vegetables in cancer survivors compared to participants with no cancer diagnosis (34). This difference might be explained by the diet score that was used to evaluate diet quality. In addition to fruits and vegetables consumption that were included in the earlier study, the diet score used in this study included additional food components to evaluate diet quality, thereby making it difficult to compare studies. Furthermore, as previously mentioned, the analyses in this thesis were based on the Lifelines diet score which follows the Dutch dietary guidelines and might have substantial differences to other diet scores assessed and calculated in different populations (35,36).

In chapter 4, a longitudinal analysis with a median follow-up of eight years, was performed to evaluate if different diet indices were associated with a lower risk of gastrointestinal cancer developments. None of the included indices were associated with a reduced cancer risk. Only a higher diet quality assessed by
the highest quintile of the American Cancer Society (ACS) index, was associated to a significant risk reduction of colorectal cancer. Finally, in chapter 3 where machine learning algorithms were used to evaluate the predictive performance of health behaviours to incident cancer cases, the addition of food components separately to the models, showed a relevant contribution of savoury and ready products to the predictive performance of health behaviours on incident prostate and gastrointestinal cancers.

Although this was a consistent finding in those models, the results still need to be validated in external population-based cohorts. The results from chapters 3 and 4 are in line with the evidence from a recent meta-analysis (5) showing that a higher diet quality is associated with a lower cancer risk. Nevertheless, this meta-analysis also mentioned that evidence is still inconclusive and that more evidence is needed in different populations to make more consistent recommendations on diet quality. Hence, evidence at this moment is not sufficient to develop dietary guidelines for gastrointestinal cancers. The results from this thesis add to evidence from a specific population from Northern Europe. Further studies on this topic may bring more certainty in the short term. This may require a longer follow up, evaluation of dietary patterns not only by food frequency questionnaires but also by using new technological advantages such as web app measures(37), that could bring less bias in the assessments.

**Physical activity**

Recent evidence suggests that leisure time physical activity associates with a reduced risk of certain cancer types such as breast, colon, rectal or lung (38), but also relates to better health during the cancer continuum, especially in cancer survivors (7). In Lifelines, physical activity was measured through validated questionnaires (39). The assessment of commuting and leisure time was transformed to moderate-to-vigorous physical activity (MVPA) in minutes per week. Previous research observed that less than 40% of Dutch population are engaged in moderate to vigorous physical activity as recommended in the Dutch guidelines (40). For incident cancers, when including physical activity as a predictor in the models at chapter 3, no substantial contribution of physical activity was observed in the predictive performance of any the models. The findings in chapter 3 can be explained partly because of the inclusion of a relatively homogeneous population that share similar commuting and leisure activities during the day in the North of The Netherlands (39).

**Chapter 5** showed an opposite trend in cancer survivors: they are more likely to be physically active compared to participants with no previous cancer diagnosis. This means that they tend to follow the recommended physical activity guidelines for Dutch people. The evaluation of physical activity as a classifier of cancer survivors compared to those without cancer by using machine learning algorithms in...
**Chapter 6** did not show it as a relevant factor for classifying cancer survivors from those without a cancer diagnosis. Some studies in population-based cohorts are consistent with the findings in **Chapter 5**, reporting cancer survivors being more physically active (27,30); while other studies reported that cancer survivors are less physically active (31). These contrasting results from the literature may be explained by several factors: i) reporting bias, in which participants could under or over report their real physical activity influencing the results of the studies (39), ii) lack of information about the severity of treatment, because more intense treatment could have reduced the physical strength of the participant, thereby altering their possibilities to be physically active, iii) time since diagnosis: longer survival may imply a better recovery and higher chances to be physically active (41,42), or iv) cultural differences, simply because some populations are used to different daily routines and this could influence their ability to be physically active(40). In addition, physical activity data was collected by using questionnaires that, though they are valid, are susceptible for recall bias and may not reflect the complete physical activity from the participants or change in a short period of time, reducing the effect in the analyses (39). Therefore other methods such as the inclusion of technological tools like Fitbit or smart watch (43) could reduce bias in the assessment. Finally, it is worth to mention that physical activity is not at optimal levels within the Dutch population (40), and should continue being promoted among cancer survivors, but also in the general population (7,38).

**Body mass index**

As a proxy measure of general health, normal BMI has been associated to good general health and a high BMI has been shown to be associated with different types of cancer (4). The adjusted analysis from **Chapter 5** showed that cancer survivors had a lower-normal BMI compared to those without a cancer diagnosis. This may be explained by the natural course of the disease in which the diagnosis and the treatment reduce their weight. However, after surviving cancer, these participants continued to have a lower-normal BMI. This contrasted with other studies where a higher BMI was shown in cancer survivors (30,31). A possible explanation for these distinct findings is that analyses in **Chapter 5** did not stratify by type of cancer which may have had an impact in the BMI composition of the participants, also the cross-sectional design does not allow to explore changes in BMI before and after cancer diagnosis and treatment. In **Chapter 6**, the analyses from **Chapter 5** were further evaluated by using both linear and machine learning algorithms to explore differences in health behaviours of cancer survivors compared to those without a history of cancer. The findings from **Chapter 6** showed that BMI was consistently the second most important contributor to the classification performance of the linear and non-linear
models, however not relevant enough to classify cancer survivors from those without a cancer diagnosis. Possible explanations for these findings might be consistent with the explanation formulated in chapter 5: lack of stratification by cancer type. However, the major limitation was the cross-sectional design of the study that did not allow evaluation of BMI during the natural course of the disease and treatment in order to observe potential changes in BMI of the participants. Hence future longitudinal analyses may shed light on the effect of changes in BMI during the course of the disease. In Chapter 3, BMI was included in the linear and non-linear algorithms as a predictor of all cancers, breast, prostate and gastrointestinal cancer incidence. In none of the proposed models BMI was a relevant contributor to the predictive performance of the models. This lack of predictive power of BMI is in line with the results of a recent meta-analysis (4) in which only a clear association between increased BMI and six cancer types was found: leukaemia, multiple myeloma, pancreatic, endometrial, rectal and renal. In chapter 3 these types of cancer were not evaluated separately due to the small number of people with those cancer types.

Sedentary behaviour

In this thesis, sedentary behaviour was included as a separate variable from physical activity, measured by the amount of television watching hours that a participant reported in Lifelines. The reason for a particular interest in sedentary behaviour was because this factor has been independently associated with colorectal cancer risk in a recent study and a meta-analysis (8,44). In chapter 5, sedentary behaviour was not substantially different between cancer survivors and those without cancer. Also in chapter 3, sedentary behaviour was not identified as a relevant predictor in the models. Possible explanations for these findings are that in chapter 5 specific cancer types were not separately analysed (44). In addition, the earlier mentioned study was only conducted in males (44) while in the analysis of chapter 5 both males and females were included.

In this thesis, sedentary behaviour was included in the analyses of chapters 3, 5 and 6, as a separate factor from physical activity. The observed results in those chapters were not significant. However, other studies evaluating health behaviours in cancer survivors did not have the opportunity to evaluate sedentary behaviour independently. Hence, the inclusion in this thesis of sedentary behaviour as an independent factor, may give an added value to the different approaches used in chapters 3, 5 and 6.

Methodological considerations

First, the analyses performed in this thesis were conducted using the data from Lifelines, the large
population-based cohort from the North of The Netherlands, that includes 147,000 adult Dutch participants. Due to the low selection bias and the representativeness of the cohort (45,46), the findings based on the analyses of the Lifelines data could be generalized to the population in The Netherlands (45). It is crucial to mention that in the baseline assessment of Lifelines there were more than 110,000 participants with complete evaluation (no missing values) of health behaviours as presented in chapter 5. In chapter 3, when the whole cohort including missing data was used, the values were then imputed using multiple imputation chained equations (47).

Second, in chapter 2, an algorithm was proposed to homogenise self-reported cancers from the participants in Lifelines, that was originally collected from free text (handwritten). The implementation of this algorithm allowed self-reported cancers in chapter 2 to be consistent with the international classification norms (43). However, the decisions taken during the creation of this algorithm may have led to potential misclassifications. For instance, the terms used to make such decisions were adapted from the free text reports of participants and those are susceptible to misreporting. These rules and decisions made for this algorithm for this specific population need to be checked if other researchers adapt it to have more consistent evaluation of their self-reported cancer data in agreement to internationally accepted criteria. In addition, the validity of self-reported cancer derived from this algorithm showed moderate to high agreement, when it was evaluated by the linkage with PALGA, which has a nationwide coverage of the pathology reports in The Netherlands (19).

Third, the cross-sectional design of the studies in this thesis regarding cancer survivors in chapters 5, 6 and 7, though informative, raise some questions, such as the level of exposure in time to every health behaviour and the potential changes that occurred in the participants after cancer diagnosis and after treatment of their specific cancer. Those may need to be better answered in longitudinal studies to properly assess the effect of long-term exposure of the health behaviours and give a better insight into how cancer survivors evolve in their lifestyle after diagnosis.

Fourth, the longitudinal studies in chapters 3 and 4 are proxies for evaluating the impact of health behaviours on incident cancers. Linear and non-linear approaches were compared to explore the impact of statistical approaches on the associations between health behaviours and incident cancers. This was a point of interest because there is an increasing use of machine learning approaches claiming higher classification/prediction performance over traditional linear methods. On the other hand, referring to clinical prediction models, recent literature states that there is no sufficient evidence to claim a better performance of machine learning algorithms over traditional methods (48). Furthermore, previous
literature shows that in scenarios where machine learning algorithms had a better performance than commonly used statistics, there is a very high risk of bias in the validation procedures being present, in addition to other limitations (i.e. relatively small sample size, limited number of predictors or limited information about handling missing data) (48). The proposed solution was developing the models in this thesis adhering to new reporting guidelines to minimize the risk of bias during model evaluation (49). These specific reporting guidelines were used in chapters 3 and 6 to improve the methodological approach and reduce bias in the prediction-classification models (49,50). The results suggested no substantial differences in the overall results comparing linear and non-linear approaches in the prediction/classification performance of the cancer outcomes. This might be explained by the adherence to reporting guidelines, which might have decreased bias in the analyses. As stated in previous literature (48), the models with high risk of bias had higher difference in their performances compared to those with low risk of bias. Hence, it could be that due to the reduced bias, the models from the aforementioned chapters might have produced moderate predictive performance. Nevertheless, both linear and non-linear models should continue being used in future research (following the established guidelines) to evaluate the associations between lifestyle behaviours and cancer, as they might have different outcomes for other populations that are not as homogeneous as the Lifelines cohort.

**Clinical implications**

In The Netherlands, cancer is the leading cause of mortality (51,52) and mainly due to aging, the number of incident cancers and cancer survivors continue to increase steadily. Evidence from this thesis concerning the role of health behaviours to the development of cancer, or their presence in cancer survivors, resulted in the following points that are in line with the existing literature and give some hints to be further studied: i) the relation between smoking, diet and alcohol intake and incident cancers, ii) the still inconsistent evidence about the relation between physical activity, sedentary behaviour and BMI with the occurrence of cancer, and iii) the role of those health behaviours in cancer survivors.

In part 2 of this thesis, the role that smoking and diet have on the incidence of gastrointestinal cancers is consistent in both chapters. Chapter 4 provides additional evidence in terms of dietary intake to reduce the incidence of gastrointestinal cancers, and shows a potential benefit of evaluating and adhering to recommendations to dietary guidelines developed by the ACS, thus promoting a healthier diet in Dutch population. Perhaps this could help to improve or re-evaluate currently established guidelines and incorporate ACS recommendations. In addition, it is worth to evaluate if those recommendations are known and promoted for the general population. The rest of the health behaviours, namely BMI,
physical activity, and sedentary behaviour did not show an association with incident cancer cases, this may be due to the relatively short follow up time. This opens the opportunity to re-evaluate these conditions in the Lifelines cohort in the future, as the data collection will be continued after finalizing this thesis. Hence, new evidence could emerge with the longer follow up. In addition, in chapter 3, levels of savoury and ready product intake consistently contributed to the improved predictive performance of the models, especially for predicting prostate cancer. Therefore, this should be further studied, and if the associations of savoury and ready product intake to incident cancers will be clearly addressed in other studies, then those might be considered in clinical advice. In part 3, chapter 7 provides evidence about role of the poor diet quality in cancer survivors and in the general population, showing that the recommended number of fruits and vegetables is not achieved by both groups. This could be addressed by launching an information campaign first focusing on cancer survivors, but also on the general population in which a better diet quality based on positive food groups should be promoted. These positive food groups are: fruits, vegetables, whole grain products, legumes & nuts, fish, oils & soft margarines, unsweetened dairy, coffee and tea. Smoking cessation programs are efforts that have been adapted worldwide, in general to reduce the risk of respiratory diseases, and specifically to reduce lung cancer incidence (1,53). The evidence from chapter 5 suggests that even cancer survivors, especially females with prolonged survival are showing a health risk smoking behaviour, and this behaviour should also be diminished to avoid a potential cancer recurrence. Reducing alcohol consumption is another factor that should be addressed in the recommendations for the general population, but especially for cancer survivors, because only male cancer survivors showed a tendency in reducing alcohol intake, though it was not significant. Clinicians might find other strategies to encourage improving lifestyle with a clearer focus on the afore mentioned factors.

**Future perspectives in research**

In this thesis, the role of health behaviours in incident cancer and their presence in cancer survivors was evaluated in some cross-sectional studies, or studies including a relatively short follow up period. To further explore that association, a longitudinal analysis with longer follow up time and repeated measurements should be conducted, thus better showing the effects of long-term exposure of health behaviours on cancer incidence and recurrences.

In this thesis separate analyses were performed only for incident gastrointestinal cancers because of the direct impact that health behaviours have on those types of cancer (5,24), but the same should be done for other cancers, such as breast or prostate cancer. Specific factors related to cancer must be included
in the future classification/prediction models, for instance, genetic variants such as BRCA1/BRCA2 for breast cancer or specific biomarkers such as prostate antigen. This might help to make better classifications and anticipate a potential risk for developing a malignancy. Also, some studies have considered clusters of health behaviours in order to assess their combined effect on incident cancers. This was beyond the scope of this thesis but may also be an area for exploring in future studies. In addition, it may be important to include in the analysis other comorbidities such as diabetes or cardiovascular disease that might have an impact on the general health of the individual together with cancer.

The studies about health behaviours in cancer survivors in this thesis are the largest in The Netherlands, to the best of our knowledge. Nevertheless, the chapters in part 3 used mainly self-reported cancers as outcomes, therefore, future evaluations regarding cancer survivors could better use the precise diagnoses from PALGA to be more precise on the analysis. Moreover, the analysis of cancer survivors with specific cancer types is warranted, such as breast, prostate or skin cancer, which were not explored in this thesis. This might add more detailed information about how the specific diagnoses relate to the health behaviours of cancer survivors.

The inclusion of sedentary behaviour as a separate variable apart of overall physical activity as was done in this thesis for chapters 3, 4, 5 and 6, would provide relevant information for future investigations if longer follow up is available and for certain types of cancer like colon, rectal and gastrointestinal. Finally, further research would benefit from having a more detailed evaluation of health behaviours, not only by questionnaires, and repeated measurements. Further evidence should also be supported, scrutinized and validated with robust research in other parts of the world. Geographical, socioeconomic and cultural differences impact the results found in this thesis.

Conclusions

From the first part of this thesis, it can be concluded that self-reported cancer in the Lifelines cohort has high positive predictive value and moderate sensitivity. The high positive predictive value is an indicative of high quality of the PALGA registration. When a cancer is self-reported it is most likely to be present in the pathology records. The underreports of cancer in Lifelines by the participants were the cause for a moderate sensitivity. In addition, self-reported cancer from Lifelines can be used in future research, except for skin and cervical cancers since those types were the main source of underreported cases.

From the second part of this thesis on incident cancer cases comprising chapters 3 and 4; it can be
concluded that the performance of machine learning algorithms for predicting incident cancer cases in the Lifelines cohort is moderate to low, and did not outperform traditional approaches, this could be due to the included variables did not show relevant differences but also the relatively short follow-up. Another conclusion from this part of the thesis is that interventions to improve diet quality could be based on the ACS score, as this score was the one showing better diet benefits for the general population included in the Lifelines cohort. Lastly from this part, the remaining health behaviours: physical activity, sedentary behaviour, and BMI did not show relevant influence on incident cancer cases.

From the third part of this thesis on cancer survivors, comprising chapters 5, 6 and 7; it can be concluded that health behaviours were not at the expected levels neither in cancer survivors, nor in the whole cohort. It was observed that lifestyle behaviour in cancer survivors needs a lot of improvement, however they have a lower-to-normal BMI and are more physically active than the rest of the cohort. Targeted interventions should be prioritized in females and older cancer survivors to promote smoking cessation and to improve their diet quality. Finally, frequent evaluation of health behaviours is important and would allow the prompt actions that decrease risks and improve lifestyle thus protecting against cancer.
concluded that the performance of machine learning algorithms for predicting incident cancer cases in the Lifelines cohort is moderate to low, and did not outperform traditional approaches, this could be due to the included variables did not show relevant differences but also the relatively short follow-up.

Another conclusion from this part of the thesis is that interventions to improve diet quality could be based on the ACS score, as this score was the one showing better diet benefits for the general population included in the Lifelines cohort. Lastly from this part, the remaining health behaviours: physical activity, sedentary behaviour, and BMI did not show relevant influence on incident cancer cases.

From the third part of this thesis on cancer survivors, comprising chapters 5, 6, and 7; it can be concluded that health behaviours were not at the expected levels neither in cancer survivors, nor in the whole cohort. It was observed that lifestyle behaviour in cancer survivors needs a lot of improvement, however they have a lower to normal BMI and are more physically active than the rest of the cohort. Targeted interventions should be prioritized in females and older cancer survivors to promote smoking cessation and to improve their diet quality. Finally, frequent evaluation of health behaviours is important and would allow the prompt actions that decrease risks and improve lifestyle thus protecting against cancer.

References


29. Ferrari P, Jenab M, Norat T, Moskal A, Slimani N, Olsen A, et al. Lifetime and baseline alcohol intake and risk of colon and rectal cancers in the European Prospective Investigation into Cancer and


44. Cao Y, Keum NN, Chan AT, Fuchs CS, Wu K, Giovannucci EL. Television watching and risk of


