Better decision making in cancer:

screening tests and prediction models

Maaike Alblas

Better decision making in cancer: screening tests and prediction models
Thesis, Erasmus MC, University Medical Center Rotterdam

The studies described in this thesis received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 634570.

Financial support for the printing of this thesis was provided by the Department of Public Health of the Erasmus MC.

Cover design: Remco Wetzels
Layout and printing by: ProefschriftMaken || www.proefschriftmaken.nl

Better Decision Making in Cancer: Screening tests and prediction models

Verbeterde besliskunde bij kanker: screeningsonderzoeken en predictie modellen

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
prof. dr. A.L. Bredenoord
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 15 december 2021 om 10.30 uur

door

Maatje Marije Alblas - van der Meer
geboren in Harlingen

Erasmus University Rotterdam

Erasmus

Promotie commissie

Promotor:	prof. dr. E.W. Steyerberg
Overige leden:	prof. ir. dr. M.K. Schmidt
	prof. dr. C. Verhoef
	dr. I. Lansdorp - Vogelaar
Copromotor:	dr. N. van Leeuwen

Paranimfen

S.A. Dijkland

K.A. Foks

CONTENTS

Part I

General introduction

Chapter 1

General introduction

GENERAL INTRODUCTION

Over the past decades, cancer has become one of the most important causes of death worldwide. It is estimated that there were 18.1 million new cancer cases and 9.6 million deaths due to cancer in 2018. Lung cancer and breast cancer are the most frequently occurring cancers (12% of all cancers), and lung cancer resulted in the highest number of cancer deaths. [1, 2] The resulting global cancer burden has increased and cancer incidence and mortality are expected to spike even further in the future.[1, 2] This trend does not only affect individuals' quality of life and future perspectives, but also challenges global resources in terms of health care capacity, and puts a strain on available health care budgets[2]. Since 2000, total health care expenditure has outgrown the gross domestic product (GDP) by approximately 1.2% on a yearly base.[3] More specifically, the global annual economic burden of cancer was estimated at 1.16 trillion US dollar in 2010, and has continued to rise ever since.[4] Given these worrying statistics on the burden of cancer for both individuals and society, there is a need for action to change cancers' course, specifically focusing on prevention. Prevention can be divided into three categories: primary, secondary and tertiary prevention. Primary prevention is focused on preventing cancer before it occurs, for example by stimulating healthy behavior (e.g. campaigns to prevent young adults from smoking). Secondary prevention aims to detect cancer in an early stage to reduce the impact of cancer and improve patient outcomes. For example, early stage asymptomatic breast cancer that is detected during mammography screening usually has better treatment options and survival rates compared to symptomatic breast cancer with possible metastases. Tertiary prevention is aimed at managing cancer after the diagnosis by treatment and rehabilitation. All three types of prevention can contribute to reducing the burden of cancer. Furthermore, risk prediction could help in targeting treatment and management for individuals. For example, a risk prediction model that can identify individuals with a high risk of developing breast cancer, can target more frequent mammography's for those individuals and less frequent for low-risk individuals. In this thesis, screening and prediction will be studied for specific cancer types to evaluate ethical and cost-effectiveness of cancer screening and to develop and validate prediction models.

1.1 Screening

Cancer screening entails testing of asymptomatic individuals for abnormalities that indicate the presence of pre-cancer or (early stage) cancer. Because of the timing of screening (before individuals become symptomatic), it often results in better patient outcomes in terms of mortality, morbidity and treatment possibilities. For example, if a precursor lesion is detected before it develops into clinical cancer, curative treatment strategies might still be an option. Furthermore, the burden of treatment for individuals and society are lower compared to treatment of more advanced cancer stages. Treatment of early stage cancer results in lower (co) morbidity and lower complication rates, leading to lower health care expenditure.

Given the age-specific risks of cancer, screening is usually targeted at particular age groups to limit harms (i.e. overdiagnosis and unnecessary treatment) and to optimize its effectiveness. A screening program is considered effective if it reduces morbidity or mortality. Several cancer-screening programs have been developed over the past decades, and with some implemented worldwide. Examples of successfully implemented screening programs for the general population include mammography screening for breast cancer, pap smear for cervical cancer and stool testing for colorectal cancer. Since the implementation of these screening programs, the incidence of most of these types of cancer has dropped notably.[5-7] Disease-specific mortality rates have also decreased. Mammography screening reduced the breast cancer-specific mortality with 13-17%.[8, 9]

The health benefits of a screening program have a price tag. Obviously, every screening program costs money, which varies per screening program. For example, mammography screening in the Netherlands costs approximately €67 per person. On average, one million women participate in this test yearly, which sets the total costs of mammography screening in the Netherlands at 67 million euros per year.[10] Although the costs of a screening program are rather straightforward, it is unwise to make decisions regarding the implementation of these screening programs based on costs alone. Hence, one should combine the effectiveness in terms of health benefits and costs to determine its cost-effectiveness. Cost-effectiveness is defined as the difference in costs divided by the difference in life years (Lys) or quality-adjusted life years (QALYs). These (quality-adjusted) life years can be compared to no screening (average cost-effectiveness ratio, ACER), or compared to other screening strategies (incremental cost-effectiveness ratio, ICER). The optimal strategy is defined as the strategy that has the highest benefit with an ICER below the willingness-to-pay (WTP) threshold, which can vary between countries.[11] The WTP threshold depends on local policies, and is influenced by the incidence and mortality of a cancer type. For example, the United Kingdom uses a threshold of £20,000-£30,000 (approximately €24,000-€40,000) per QALY, while in the Netherlands this ranges between €20,000 per QALY for diseases with a

low burden, and €80,000 per QALY for diseases with a high burden.[12, 13] Both thresholds are used as directive, and decisions regarding the inclusion of a specific treatment or screening program in local health care reimbursements may differ per disease.

Cancer screening programs are not solely targeted at (seemingly) healthy individuals. The discovery of genetic mutations that are responsible for an increased risk of cancer has enabled specific targeting of high-risk individuals, the so-called (epi)genetic risk-tailored cancer screening. An example is reflex testing of patients diagnosed with colorectal cancer to determine the presence of Lynch Syndrome. Lynch syndrome is caused by a DNA mutation, which substantially increases the risk of colorectal, endometrial, and ovarian cancer.[14] Therefore, some guidelines recommend testing every patient diagnosed with colorectal cancer for Lynch Syndrome.[15-17] If Lynch Syndrome is present, the first-degree relatives of these patients can also be tested for Lynch Syndrome, as this mutation is hereditary. Information on the presence of Lynch Syndrome can aid in further preventive strategies, such as offering prophylactic surgery and more frequent cancer screening of relatives as part of secondary prevention.

Although these new technological advances create new possibilities for risk-tailored cancer screening, it also requires individuals whom participate in such screening programs to process a lot of complex information. Several studies have shown that people are often incapable of processing large amounts of information, especially if this information includes benefits, risks and its trade-offs.[18-20] This might compromise the principle of patient autonomy regarding the decision to participate in risk-tailored screening programs, as it unclear if an individual is able to make a well-informed decision.

1.2 Prediction

Prediction models are empirical tools to estimate the probability of an event of interest based on multiple clinical or other variables, which can be used to optimize medical decision-making in the era of personalized medicine.[21] Prediction models become increasingly relevant in the era of personalized medicine. These models are mostly based on multivariable regression analyses, which can be presented with a single risk score. Models can also estimate the risk over a certain amount of time (i.e. 1-, 2- and 5-year risks).

In the field of cancer, these models can be used to estimate the probability of diagnosing cancer, or to estimate the probability of progression from a premalignant lesion to clinical cancer. Prediction models provide estimates for risk of cancer for individual patients, which makes their use in clinical practice preferable over risks based on the general population. For instance, a risk estimate for cancer based on specific patient characteristics will be more

relevant for the patient and physician. The presentation format of prediction models varies and can include a risk chart, nomogram of a web-based interactive tool. An example of an interactive tool to predict the chance of regression of hepatocellular adenomas that are found in the liver is shown in figure 1,and includes the 1 and 2-year chance of progression for a specific patient.[22] The use of prediction models in clinical practice has gained popularity.[23] Although this should improve the practice of personalized medicine, the use of poor quality prediction models in clinical practice might lead to harmful decisions because risk estimates might be too high or too low (poor calibration or discrimination). [24]

Predictors		**Value**
Diameter at diagnosis	*[mm]*	70
Date diagnosis	*[dd-mm-yyyy]*	20-09-2017
Diameter at first follow-up	*[mm]*	60
Date first follow-up	*[dd-mm-yyyy]*	01-03-2018
Subtype	*[0=H-HCA | 1=I-HCA | 2=U-HCA]*	1

Predicted chance of regression to <5cm (%)	
1 year after diagnosis	**7**
2 years after diagnosis	**29**

Figure 1 – Example of a tool to predict the change of regression of hepatocellular adenomas[22]

The performance of prediction models can be assessed in terms of discrimination and calibration. The discriminative ability of a model indicates how well the model can distinguish between patients with and without the outcome of interest (e.g. a cancer diagnosis). Calibration refers to the comparison of predicted risks that are estimated by the model versus observed risks in the population.[21, 25] It is important to determine the quality of prediction models before implementing them in clinical practice. Every prediction model should be validated with appropriate methods in the development cohort (internal validation) and preferable also in an external cohort that differs from the development cohort (external validation). External validation may support claims of generalizability to new settings.[25, 26] Ideally, such validation should be followed by an impact analysis to determine the effect of the predictions provided by the model on patient outcomes. [23, 26] Several studies have shown that the quality of prediction models varies and that external validation is often lacking.[27, 28]

A number of guidelines have been developed to improve the reporting of prediction models to enable proper assessment of model quality.[29, 30] Furthermore, recent developments in the field of 'omics' (e.g. information from the genome, epigenome, and microbiome) might contribute to the increasing the quality of prediction models, as omics can add new information (that is currently lacking) to models with poor discriminatory ability. For example, several studies are exploring the possibility of improving cancer risk prediction models with information from the epigenome.[31] [32] The Female Cancer Prediction Using Cervical Omics to Individualize Screening and Prevention (FORECEE) project is currently studying

the possibility of developing a risk prediction model for breast, ovarian, endometrial and cervical cancer that incorporates information from the genome, epigenome, metagenome and epidemiological variables (figure 2).[33]

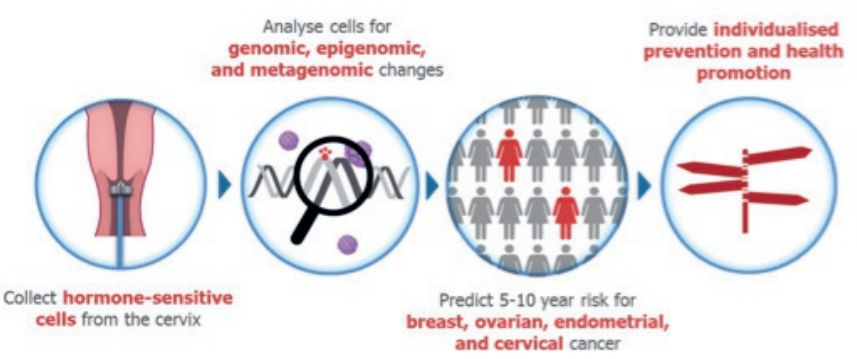

Figure 2 - Developing the Women's Cancer Risk Identification Tests. Source: FORECEE 2015

1.3 Application of screening and prediction

Medical decision-making refers to the process of decision making of both the patient and the physician. Either way, medical decisions often include complex trade-offs of benefits and harms, have a relatively high level of uncertainty, and are preference sensitive[34]. For example, women have to make a trade-off between the reduction in morbidity and mortality of breast cancer, and the risk of overdiagnosis and overtreatment, before consenting to participate in mammography screening.[35] These benefits and harms usually have different dimensions and scales, which complicates their interpretation and comparison. Moreover, there are also negative effects of screening on quality of life that are difficult to express in numbers. People might feel stressed when receiving an invitation for screening and worry about the possibility of having cancer. This overwhelming amount of information should be integrated into the decision of the patient and the physician; a seemingly difficult task. Hence, there is a need for guidance and support.

Outcomes of screening and prediction research can provide additional information to physicians and patients, and enable them make more individualized and personalized decisions. Specific risk scores for a high-risk population with a genetic mutation could also feature a higher level of information compared to a general population-based risk score. Female carriers of the BRCA 1 and BRCA 2 have a lifetime risk of breast cancer of respectively 72% and 69%, while women in the general population have a lifetime breast cancer risk of 12%.[36] This substantial difference in risk estimates may influence a women's decision for treatment strategies; i.e. BRCA 1 and BRCA 2 mutation carriers might undergo prophylactic

mastectomy to eliminate their risk of breast cancer if they are aware of their specific (high) risk.

Both screening and prediction will be addressed in this thesis, and include the development of prediction models for specific cancer types (application in pancreas, liver and skin cancer) and a simulation model to estimate the effectiveness of prophylactic surgery in patients with a high risk of a endometrial cancer. The incidence and mortality rates of these cancer types are shown in Table 1.[1] Furthermore, ethical aspects on risk-stratified cancer screening are studied.

Table 1. Global incidence and mortality of cancer types[1] addressed in this thesis

	Incidence (%)	Mortality[a] (%)
Pancreas cancer	2.5	4.5
Liver cancer	4.7	8.2
Skin cancer		
Melanoma	1.6	0.6
Non-melanoma	5.8	0.7
Endometrial cancer	2.1	0.9

[a] Mortality rates are shown as the average risk of death in the global population in 2018

1.3.1 Risk prediction and prophylactic surgery in endometrial cancer

Endometrial cancer is the most common type of gynecologic cancer in developed countries, and its incidence continues to increase.[1, 37] This increase might be related to the rising body mass index (BMI) of individuals, as adiposity is the strongest predictor for developing endometrial cancer[38-41]. Several prediction models have been developed that combine predictors for endometrial cancer, including BMI, to generate risk predictions for healthy women and women with postmenopausal vaginal bleeding.[42-46] However, an overview of these models and their methodological quality is lacking, which complicates their use in clinical practice.

Besides epidemiological risk factors such as BMI, it is known that genetic mutations, such as Lynch Syndrome, increases the risk of endometrial cancer. Women with Lynch Syndrome have a 40-60% lifetime risk of developing endometrial cancer[14]. Current guidelines indicate that prophylactic hysterectomy is recommended for patients with Lynch Syndrome to eliminate their risk of endometrial cancer[47]. However, the cost-effectiveness of this preventive strategy is unknown and the optimal age range to perform prophylactic hysterectomy is yet to be determined.

1.3.2 Prediction of progression of premalignant cancer to clinical cancer

In pancreas cancer, liver cancer and skin cancer, premalignant stages of cancer can be found in the form of neoplastic cysts, hepatocellular adenomas, and actinic keratosis, respectively. These premalignant stages are usually an incidental finding during routine examination, e.g. hepatocellular adenomas are often found during surveillance in women using estrogen-containing hormonal contraceptives. Although these premalignant stages differ in type and frequency of occurrence, they share the risk of turning into clinical cancer.[48-51] This warrants for surveillance of these premalignant stages, as is suggested in current guidelines. The probabilities that these premalignancies turn into cancer are however rather low. Furthermore, some of these premalignancies might regress over time and disappear on their own.[52] Given these disease-related features and the fact that the negative effects of surveillance might outweigh its benefits, it is questionable whether every patient should receive surveillance. Prediction tools might be helpful in identifying patients at low risk of developing clinical cancer and may therefore aid in selecting patients that do not require surveillance.

1.4 Aim and outline of this thesis

The overall aim of this thesis is to evaluate ethics, cost-effectiveness, and prediction modeling in the context of cancer screening and treatment. The following specific research questions will be addressed:

1. How can we promote autonomous choices in informed consent procedures regarding participation in epigenetic risk-tailored cancer screening?
2. What is the cost-effectiveness of offering prophylactic hysterectomy to first-degree relatives with Lynch Syndrome of patients with colorectal cancer?
3. What is the quality of risk prediction models for endometrial cancer in the general population?
4. How can we obtain accurate and valid individual estimates of progression from premalignant to malignant cancer based on prediction models to guide decision making?

Part II of this thesis concerns cancer screening and aims to answer research questions 1 and 2. **Chapter 2** discusses a framework to support autonomous choices in epigenetic risk-tailored cancer screening. **Chapter 3** presents a cost-effectiveness study on offering prophylactic hysterectomy in first-degree female relatives with Lynch Syndrome.

Part III of this thesis focuses on prediction modeling in cancer and aims to answer research questions 3 and 4. **Chapter 4** provides an overview of risk prediction models for endometrial

cancer in the general population and discusses their methodological quality. **Chapters 5, 6, and 7** describe the development and validation of prediction models for hepatocellular adenoma, intraductal papillary mucinous neoplasms, and keratinocyte carcinoma, respectively.

Part IV of this thesis contains the general discussion with answers to the four research questions. These answers are compared with the existing literature, and future perspectives are discussed.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2018;68: 394-424.
2. Soerjomataram I, Lortet-Tieulent J, Parkin DM, et al. Global burden of cancer in 2008: a systematic analysis of disability-adjusted life-years in 12 world regions. The Lancet. 2012;380: 1840-1850.
3. Xu K SAKJea. Public Spending on Health: A Closer Look at GlobalTrends. World Health Organization. 2018.
4. Stewart BW, Wild CP. World Cancer Report 2014. International Agency for Research on Cancer, 2014.
5. Broeders MJM, Allgood P, Duffy SW, et al. The impact of mammography screening programmes on incidence of advanced breast cancer in Europe: a literature review. BMC cancer. 2018;18: 860-860.
6. Peto J, Gilham C, Fletcher O, Matthews FE. The cervical cancer epidemic that screening has prevented in the UK. The Lancet. 2004;364: 249-256.
7. Hewitson P, Glasziou P Fau - Watson E, Watson E Fau - Towler B, Towler B Fau - Irwig L, Irwig L. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update.
8. Gotzsche PC, Jorgensen KJ. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2013: CD001877.
9. Independent UKPoBCS. The benefits and harms of breast cancer screening: an independent review. Lancet. 2012;380: 1778-1786.
10. Integraal kankercentrum Nederland RvVeM. Monitor bevolkingsonderzoek borstkanker 2017-2018, 2018.
11. Mark DH. Visualizing cost-effectiveness analysis. Jama. 2002;287: 2428-2429.
12. Nice U. Guide to the methods of technology appraisal: National Institute for Health and Clinical Excellence (NICE) London, UK, 2008.
13. Jacqueline Zwaap SK, Caroline van der Meijden, Paula Staal, Luuk van der Heiden. Cost-effectiveness in practice. In: Nederland Z, editor, 2015.
14. Bonadona V, Bonaiti B, Olschwang S, et al. Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. Jama. 2011;305: 2304-2310.
15. Lindor NM, Petersen GM, Hadley DW, et al. Recommendations for the Care of Individuals With an Inherited Predisposition to Lynch SyndromeA Systematic Review. Jama. 2006;296: 1507-1517.
16. Vasen HF, Blanco I, Aktan-Collan K, et al. Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. Gut. 2013;62: 812-823.
17. Stoffel EM, Mangu PB, Gruber SB, et al. Hereditary colorectal cancer syndromes: American Society of Clinical Oncology Clinical Practice Guideline endorsement of the familial risk-colorectal cancer: European Society for Medical Oncology Clinical Practice Guidelines. J Clin Oncol. 2015;33: 209-217.
18. Ubel PA. Beyond costs and benefits: understanding how patients make health care decisions.

19. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. Science. 1974;185: 1124-1131.
20. Wegwarth OA-Ohoo, Widschwendter M, Cibula D, et al. What do European women know about their female cancer risks and cancer screening? A cross-sectional online intervention survey in five European countries.
21. Steyerberg EW. Clinical Prediction Models. Springer-Verlag New York, 2009.
22. Klompenhouwer AJ, Alblas M, van Rosmalen BV, et al. Development and Validation of a Model to Predict Regression of Large Size Hepatocellular Adenoma. Am J Gastroenterol. 2019;114: 1292-1298.
23. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med. 2006;144: 201-209.
24. Van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. BMC Medicine. 2019;17: 230.
25. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35: 1925-1931.
26. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98: 691.
27. Bouwmeester W, Zuithoff NP, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS medicine. 2012;9: 1-12.
28. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. Breast Cancer Res Treat. 2012;132: 365-377.
29. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD StatementThe TRIPOD Statement. Annals of Internal Medicine. 2015;162: 55-63.
30. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS medicine. 2013;10: e1001381-e1001381.
31. FORECEE Female Caner Prediction Using Cervical Omics to Individualise Screening and Prevention. Available from URL: http://www.forecee.eu/ [accessed November 20 2017, 2017].
32. Appeal TE. The Eve Appeal: Predicting-Risk of Ovarian-Malignancy Improved-Screening and Early detection (PROMISE). Available from URL: https://eveappeal.org.uk/our-research/our-research-programmes/promise-2016/ [accessed 2020/02/26, 2020].
33. Widschwendter M, Jones A, Evans I, et al. Epigenome-based cancer risk prediction: rationale, opportunities and challenges. Nat Rev Clin Oncol. 2018.
34. Hunink MGM, Weinstein MC, Wittenberg E, et al. Decision Making in Health and Medicine: Integrating Evidence and Values. 2 ed. Cambridge: Cambridge University Press, 2014.
35. Løberg M, Lousdal ML, Bretthauer M, Kalager M. Benefits and harms of mammography screening. Breast cancer research : BCR. 2015;17: 63-63.
36. Kuchenbaecker KB, Hopper JL, Barnes DR, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. Jama. 2017;317: 2402-2416.

37. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur J Cancer. 2013;49: 1374-1403.
38. Kyrgiou M, Kalliala I, Markozannes G, et al. Adiposity and cancer at major anatomical sites: umbrella review of the literature. BMJ (Clinical research ed). 2017;356: j477.
39. Dixon JB. The effect of obesity on health outcomes. Mol Cell Endocrinol. 2010;316: 104-108.
40. Collaboration NCDRF. Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants. Lancet. 2016;387: 1377-1396.
41. Ng M, Fleming T, Robinson M, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2014;384: 766-781.
42. Pfeiffer RM, Park Y, Kreimer AR, et al. Risk prediction for breast, endometrial, and ovarian cancer in white women aged 50 y or older: derivation and validation from population-based cohort studies. PLoS medicine. 2013;10: e1001492.
43. Wong AS, Cheung CW, Fung LW, Lao TT, Mol BW, Sahota DS. Development and validation of prediction models for endometrial cancer in postmenopausal bleeding. Eur J Obstet Gynecol Reprod Biol. 2016;203: 220-224.
44. Husing A, Dossus L, Ferrari P, et al. An epidemiological model for prediction of endometrial cancer risk in Europe. Eur J Epidemiol. 2016;31: 51-60.
45. Burbos N, Musonda P, Giarenis I, et al. Predicting the risk of endometrial cancer in postmenopausal women presenting with vaginal bleeding: the Norwich DEFAB risk assessment tool. Br J Cancer. 2010;102: 1201-1206.
46. Giannella L, Mfuta K, Setti T, Cerami LB, Bergamini E, Boselli F. A risk-scoring model for the prediction of endometrial cancer among symptomatic postmenopausal women with endometrial thickness > 4 mm. Biomed Res Int. 2014;2014: 130569.
47. Giardiello FM, Allen JI, Axilbund JE, et al. Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US Multi-Society Task Force on colorectal cancer. Gastroenterology. 2014;147: 502-526.
48. Stoot JH, Coelen RJ, De Jong MC, Dejong CH. Malignant transformation of hepatocellular adenomas into hepatocellular carcinomas: a systematic review including more than 1600 adenoma cases. HPB (Oxford). 2010;12: 509-522.
49. van Aalten SM, de Man RA, JN IJ, Terkivatan T. Systematic review of haemorrhage and rupture of hepatocellular adenomas. Br J Surg. 2012;99: 911-916.
50. Apalla Z, Lallas A, Sotiriou E, Lazaridou E, Ioannides D. Epidemiological trends in skin cancer. Dermatology practical & conceptual. 2017;7: 1-6.
51. Crippa S, Capurso G, Camma C, Delle Fave G, Castillo CF, Falconi M. Risk of pancreatic malignancy and mortality in branch-duct IPMNs undergoing surveillance: A systematic review and meta-analysis. Dig Liver Dis. 2016;48: 473-479.
52. Klompenhouwer AJ, Broker MEE, Thomeer MGJ, Gaspersz MP, de Man RA, JNM IJ. Retrospective study on timing of resection of hepatocellular adenoma. Br J Surg. 2017;104: 1695-1703.

Part II

Screening

Chapter 2

Autonomy Challenges in Epigenetic Risk-Stratified Cancer Screening: How Can Patient Decision Aids Support Informed Consent?

Maaike Alblas, Maartje Schermer, Yvonne Vergouwe and Ineke Bolt

Journal of Personalized Medicine. 2019

ABSTRACT

Information of an individual's epigenome can be useful in cancer screening to enable personalized decision making on participation, treatment options and further screening strategies. However, adding this information might result in complex risk predictions on multiple diseases, unsolicited findings and information on (past) environmental exposure and behavior. This complicates informed consent procedures and may impede autonomous decision-making. In this article we investigate and identify the specific features of epigenetic risk-stratified cancer screening that challenge the current informed consent doctrine. Subsequently we describe current and new informed consent models and the principle of respect for autonomy and argue for a specific informed consent model for epigenetic risk-stratified screening programs. Next, we propose a framework that guides the development of Patient Decision Aids (PDAs) to support informed consent and promote autonomous choices in the specific context of epigenetic cancer screening programs.

1. INTRODUCTION

Screening programs are important means in the prevention and early detection of several high-risk cancers [1–3]. Over the past few decades, there has been a transition from screening for a single cancer-type based on available patient characteristics, to screening for multiple cancers based on epigenetic testing [4,5]. Adding information from the epigenome to population-based cancer screening programs can eventually lead to more personalized risk predictions and treatment advice [6]. Although these new screening programs may improve the early detection and prevention of cancer, there is a drawback regarding informed consent. The screening program will include risk assessments, and therefore also difficult risk predictions that are hard to interpret for a lay person, but can also include unsolicited findings (i.e., findings that are discovered unintentionally and may have medical, psychological and social consequences) and may reveal information about environmental exposures and behavior from the past [7]. These features challenge existing informed consent procedures and call for ways to restructure and enrich informed consent procedures, with the ultimate goal that informed consent represents an autonomous choice.

Patient Decision Aids (PDAs) can support autonomous decision making [8–10]. They are proven to be effective in terms of improving knowledge and increasing patient participation [8,9,11]. However, improving knowledge and understanding does not automatically lead to autonomous choices, since an autonomous choice should be an intentional, voluntary choice that is made with sufficient understanding. Additionally, the choice should be consistent with one's values. Vos et al. discussed that the design of current PDAs may not support autonomous decision making because (1) they often utilize explicit value clarification methods (VCM), which may lead to constructed preferences that are not congruent with one's actual values, and (2) they mainly focus on deliberative processes instead of combining deliberation and intuition [12]. These findings have implications for PDAs in general, but these implications may be even more substantial for PDAs in epigenetic cancer screening given the additional challenges that are posed on informed consent in this context. Therefore, the aim of this study is to create a framework that guides the development of PDAs to support informed consent and promote autonomous choices when people face epigenetic risk-tailored cancer screening decisions regarding participating in such tests.

First, we identify specific elements of epigenetic risk-stratified cancer screening that complicate informed consent procedures. Next, we provide a brief overview of the literature on existing decision aids, informed consent models and the principle of respect for autonomy, specifically in the context of cancer screening program and argue for a specific informed consent model suitable for the context of epigenetic cancer screening programs. Subsequently, we identify different processes in PDAs to evaluate which ones sustain autonomous choices and the informed consent process in this context of epigenetic cancer

screening. Finally, we develop a framework that can serve as starting point for designing PDAs to be used within the process of obtaining informed consent, in the specific context of new epigenetic risk-stratified cancer screening programs.

2. Epigenetic Risk-Stratified Cancer Screening

New studies are currently exploring the possibility of using epigenetic markers for risk-stratification in nationwide screening programs [5,7]. The most commonly studied epigenetic changes are DNA methylation, the addition of a methyl group to specific regions of the genome. These changes affect gene expression without changing the underlying DNA sequence. Epigenetic changes are implicated in tumor development and its progression [7,13]. Unlike genetic mutations such as BRCA1/BRCA2 that are not reversible, changes in epigenetic markers are reversible and do not cause alterations in the underlying DNA sequence [14]. For instance in breast cancer, a women can inherit a BRCA1/BRCA2 mutation, which gives her a life-time elevated risk of breast cancer of 50–85%[15]. Besides preventive surgery, she cannot lower this risk by herself and the mutation will always be present in her DNA sequence. This is different in epigenetic markers, that can be influenced by heritable factors and non-heritable factors such as lifestyle, and can therefore change during a person's life [14]. If a specific epigenetic marker is responsible for an increased risk of breast cancer, this marker may be restored to its normal state with lifestyle interventions, hence lowering the risk of cancer [16].

There are studies looking into using epigenetic changes for risk assessment in risk-stratified screening programs. Risk-stratified screening means that eligibility, frequency and modality are tailored according to one's risk level [5]. Using epigenetic markers in risk stratified screening creates opportunities to incorporate advice on specific preventative actions [16]. Epigenetic markers hold valuable information on an individuals' exposure to e.g., obesity, smoking, and the use of exogenous hormones. This information from the epigenome can serve as a surrogate for more subjective data on lifestyle exposure such as questionnaire data that is prone to recall bias, and might improve the performance of cancer risk prediction models as such epigenetic markers will give more accurate risk prediction than risk assessment based on questionnaire based informed risk factors [5,7].

When epigenetic information is incorporated in nationwide screening programs, the results of these screening programs will differ from the programs that we are currently familiar with, such as breast cancer screening based on mammography. We want to illustrate this with a screening test which is under development, the FORECEE Women's cancer risk Identification (WID) test [17]. In short, the WID test aims to predict the risk of developing breast, ovarian, endometrial and cervical cancer by using (epi)genetic information that is obtained with a

pap smear [17]. The results of this test will include risk predictions on the four cancers, but may also include unsolicited findings (e.g., the risk of lung cancer). The debate regarding returning these unsolicited findings is still ongoing, and there is currently no consensus on the most appropriate disclosure policy [18]

Based on this description of epigenetic risk-stratified cancer screening programs, we can discern the following features of these type of screening tests. (I) The results of the test will not only consist of a diagnosis, but will also contain risk predictions on multiple (malignant) diseases. These risk predictions pose additional challenges on patients' ability to make autonomous choices. Several studies have shown that people are more fallible decision makers than was often assumed. They incorporate emotions and perceptions into their decision making process and, for instance, label certain numbers with affective cues such as high versus low, especially if the decision involves risk predictions [12,19]. (II) The results of the test will provide insights in specific lifestyle factors that might have contributed to an elevated risk. These insights could provoke feelings of guilt, regret or shame. These feelings might be even more substantial if one's lifestyle has caused a so-called epigenetic inheritance (i.e., one's lifestyle has caused a high risk of a certain type of cancer among its offspring). (III) The resulting risk score might be lowered by lifestyle because of the reversible nature of DNA methylation. However, changing someone's lifestyle is not as simple as it seems and people might need executive capacities to set such change in motion [20]. These capacities are likely to vary between people with a low versus high social economic position, which might lead to increasing levels of inequity [21]. (IV) The results of the test might contain unsolicited findings with large variability in severity and consequences. Although unsolicited findings are also common in other tests, the specific combination of unsolicited findings, multiple diseases and risk predictions in epigenetic cancer screening leads to an increased level of complexity.

Although these elements facilitate more personalized nationwide screening programs, they also endanger informed consent procedures. This adds to the existing problems with informed consent in tests with genetic information and testing multiple diseases simultaneously, and in current population screening programs [22]. Therefore, informed consent procedures should be reconsidered and supported by suitable tools that aid patients in making an autonomous choice about participation in epigenetic cancer screening programs

3. Autonomy and Informed Consent

Informed consent is an ethical and legal requirement in different areas of health care, such as clinical practice and scientific research [23]. In its legal form, informed consent is an authorization of a certain medical intervention [23,24]. However, this form of informed

consent might not align with its original goal to protect patients against harm and to ensure the autonomy of patients [23]. An 'autonomous choice' was described by Faden and Beauchamp as an intentional, voluntary choice that is made with understanding [23]. Informed consent is considered to be an "autonomous authorization" when a patient intentionally, voluntarily and with sufficient understanding authorizes a doctor to act. Alternative conceptions of autonomy include 'authenticity' as a necessary condition for an autonomous choice; a person should be able to identify herself with a choice and this choice should be in line with her personal values [25,26]. If this is not the case, the choice she makes might not represent an autonomous choice. We agree that alignment with one's values is an important condition, in particular in the context of epigenetic screening because preferences regarding participation might vary and participation might have substantial consequences (e.g., choices regarding preventive and treatment options e.g., mastectomy, procreation, or career). Taken together, informed consent represents an autonomous choice if the underlying choice was made intentionally, voluntarily, with sufficient understanding, and is aligned with one's personal values [23,24].

If we would apply these conditions of an autonomous choice to informed consent in epigenetic cancer screening programs, the patient should at least be informed about and understand the risks, benefits, harms associated with the test, and alternative screening options before consenting to the particular medical activity [23,24]. However, this would result in an overwhelming amount of information given the before mentioned specific elements of epigenetic cancer screening (e.g., complex. information on different risk predictions, unsolicited findings, and interplay of genes). Previous studies have shown that such large amounts of information could lead to ineffective informed consent procedures and might cause anxiety and confusion amongst patients [27,28]. In addition, people are likely to make mistakes, e.g., in assessing numbers on risk and benefits of medical interventions because of beliefs on the likelihood of a certain option and previous experiences with risk estimates [19,29]. A recent study showed that the majority of women tend to overestimate their baseline female cancer risk and have limited knowledge on the benefits and harms of screening with mammography [30]. This indicates that this problem is also present in current screening programs, but we argue that it will become even more substantial in epigenetic cancer screening programs since these programs will provide more complicated risk estimates.

We argue for an informed consent model that is suitable to the context of epigenetic risk-stratified cancer screening. Bunnik et al. describe a tiered–layered–staged model, which was originally developed for commercial personal genome testing [31,32]. In short, the tiered–layered–staged model consists of three components. The first component is tiered, which means that people can consent to different parts of the treatment or screening, depending on their preferences. For instance, in the context of personal genome testing

this could mean that patients only consent to test for diseases for which medical treatment/prevention exist (that are medically actionable), or choose to also acquire results that are not directly life-saving or medically actionable (e.g., results related to reproductive health or diseases for which no treatment is available). The second component is a layer, because the model consists of several layers of information, each layer having its own level of complexity. These layers always include a baseline layer that contains information that is necessary to make a decision with knowledge, e.g., all basic information regarding the test, including resulting risk prediction, the expected false positive/negative results, over-diagnosis and the handling of unsolicited findings. The additional layers contain more detailed and complex information and can be made available depending on the patients' preferences. The last component is staged, which means that the informed consent contains several stages in time in which information is given. The actual informed consent could contain a certain timeframe in which patients can think about their decision. Although there is currently no empirical data on the use of the tiered–layered–staged informed consent in practice, tiered informed consent is already considered best practice in biobanking in genomics- and proteomics-based research [33,34].

We suggest using the tiered–layered–staged informed consent in the context of new epigenetic risk-stratified cancer screening to overcome the challenges that are posed to traditional informed consent. The tiered–layered–staged model overcomes these problems since it organizes the information in stages and categories. However, this model mainly focusses on distributing information, instead of supporting patients in processing all provided information and relating it to their values. This processing of information is necessary because solely supplying patients with information insufficiently promotes a choice that is made with understanding and in line with their values. Next to intentionality, voluntariness and understanding, authenticity is seen as a necessary condition of the principle of autonomy, as discussed above [25]. We suggest that patients can be supported in this process by using PDAs, which can be implemented within the tiered layered staged informed consent model.

4. Patient Decision Aids

PDAs are defined as tools developed to help patients make specific, informed choices in health-related decisions that align with their own values [8,35,36]. PDAs have three goals in common: (I) to inform people about the options, risk and benefits of the intervention; (II) to stimulate active participation of the patient in the decision process; and (III) to help people consider their own values and make choices congruent with these values [36]. These goals are clearly in line with the goals of informed consent as discussed above. Decision aids can

occur in many forms, varying from an information leaflet to an interactive online tool and are used in both screening and treatment decisions.

Stimulating patients to consider their own values can be aided by a VCM, which helps patients to identify decisions that are most congruent with their values and preferences [37]. The importance of values clarification in PDAs has been emphasized by the International Patient Decision Aid Standards (IPDAS) Collaboration, who recommends that every PDA should contain a VCM. Given the goals of PDAs, values clarification is inevitable because it is assumed that patients with clear insights of their own values and preferences are more likely to make choices that are aligned with these values and preferences [37]. A VCM can be both implicit (e.g., patient thinks about his options) or explicit (e.g., patient ranks different elements of the decision based on their feeling of importance) [37].

Both implicit and explicit VCMs make use of deliberative processes, which are here defined as conscious and analytical processes [38]. These processes may lead to more value-congruent decisions and a higher level of satisfaction with the choice that people have to make, because deliberative processes align with people's expectations on how health-related decisions have to be made [38]. Deliberative processes are incorporated in VCMs assuming that people need help with the consideration of their own values and preferences [39]. However, it is not proven that solely applying deliberative processes will result in in better considerations of people's values and preferences [40]. Deliberative processes might even have a negative influence on people's natural ability to distinguish between relevant and irrelevant information, which can lead to a higher level of indecisiveness in decision making [38]. Intuitive processes stimulate the ability to separate relevant from irrelevant information. Intuitive processes are here defined as simple decision strategies that rely on less conscious, cognitive processes [38]. Such processes can improve decision making because they allow for the incorporation of feelings and emotions in the decision and the integration of large amount of information [38,41]. However, including feelings and emotions in decision making can also result in decisions that are incongruent with people's values. Further, choices may be less reproducible for people because they may be influenced by temporary moods or emotions rather than thorough reasoning [38].

Since deliberative and intuitive processes have advantages and disadvantages, it might be preferable to combines these processes. de Vries et al. showed that although it might be difficult to distinguish between deliberative and intuitive processes, they seem to have different effects on decision making in PDAs [38].

Combining deliberation and intuition might be specifically interesting in the context of epigenetic risk-stratified screening cancer screening for two reasons. The first reason is related to an important advantage of intuitive processes, which is the implicit integration of

information that enables processing large amount of information. This implicit integration can be caused by affective cues or gut feelings and can be of great value in the context of cancer screening because of the large amount of information that will be provided to patients. The second reason is related to a strength of deliberation, which is that deliberation may support patients in expressing their preferences. This is of particular interest in the context of cancer screening because the decision to participate is more preference-sensitive compared to regular treatment decisions [38]. For example, regular treatment choices such as the treatment of an infection with specific types of antibiotics have clear outcomes in terms of benefit and efficiency, and the trade-offs between risks and benefits are limited. However, the outcome of epigenetic cancer screening might be more uncertain and includes trade-offs between risks and benefits, e.g., between quantity and quality of life, and between the benefits and costs of changing lifestyle. The trade-offs that people are willing to make are likely to differ because of personal values and characteristics such as attitude toward uncertainty, values attached to peace of mind, and magnitude of ones fear of cancer [19]. This makes the decision on participating in epigenetic cancer screening more preference sensitive compared to most regular treatment decisions.

Additionally, cancer screening is usually embedded in nationwide programs for which people receive a written invitation. In principle, people do not have face-to-face with a medical professional that could help with clarifying people's values in nationwide screening programs. Although it might be possible to contact a medical professional, like one's general practitioner, this is not standard practice. This makes it even more important that patients are capable of clarifying and expressing their preferences independently. Overall, it is vital that the information provided is balanced and neutral, which will limit the possibility of steering. This can be achieved by providing both survival and death rates and by providing absolute risks rather than relative risks [42].

5. Framework

We propose a framework to guide the development of PDAs in the field of epigenetic risk-stratified cancer screening for use within the tiered layered staged informed consent model. The proposed framework incorporates a VCM and a combination of intuitive and deliberative processes (Figure 1).

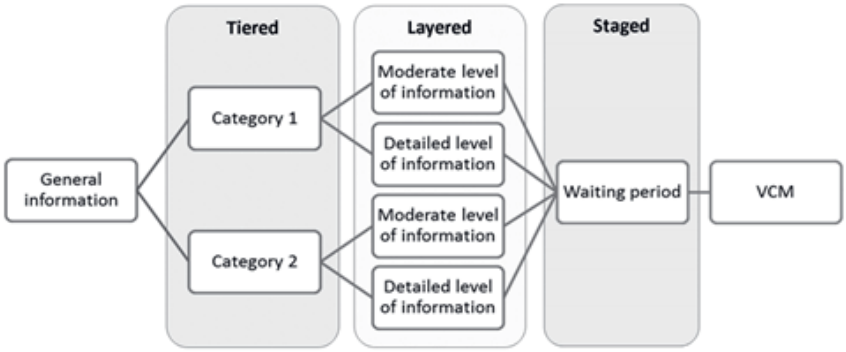

Figure 1. Framework to guide the development of Patient Decision Aids (PDAs) within the tiered layered staged informed consent model. VCM: value clarification methods.

The first part of the PDA should consist of a short section with general global information on the screening program, including the aim of the test, the target population, and the methods that will be used during the test to enable risk prediction. After the general information, the PDA should be tiered into different categories. These categories are flexible and can be chosen for each specific screening program, depending on the outcome of the screening. These categories should be chosen in such a way that they are meaningful to patients, which can be examined by focus groups or surveys. Besides these patient's preferences, ethical considerations should also be taken into account in discerning relevant categories. Categories can consists of different cancers, but they could also contain several types of unsolicited findings. A specific category, such as unsolicited findings, could then be subdivided into different categories based on condition-specific characteristics like age of onset, action ability or disease severity [32].

We want to illustrate these categories the WID test [17]. Implementation of the WID test could mean that the tiered part of the informed consent model consist of three categories. The first category contains the risk predictions for the four cancers, the second category contains unsolicited findings, which could be subdivided into medically actionable unsolicited findings (e.g., lung cancer or diabetes), and unsolicited findings which are not directly medically actionable but could be actionable in terms of procreation, career planning and other life decisions. Subsequently, the information that is presented in a particular category should be layered, as it is expected that the starting knowledge and desired level of knowledge might differ for each individual. Adopting several layers of information with increasing levels of complexity might contribute to a more personalized decision aid that suits to participants starting knowledge. The first layer of information should include all basic information regarding the test, including resulting risk prediction, the expected false positive/negative results, and over-diagnosis. The other layers can contain more in-depth information, which can consist of specific characteristics of cancers, the interplay of genes

that are used in screening with epigenetic testing, and information on that the test might reveal about lifestyle and previous exposures.

The next step in the PDA should include a waiting period, possibly with a distraction from the choice people are facing. This waiting period might improve the quality of the decision because it allows people to distance themselves from the decision, which limits the possible influence of strong feelings and emotions that can be evoked during the previous phase of the PDA [37,38]. The optimal length of a waiting period depends on the context in which the PDA is implemented. A distraction of the decision has been shown to improve intuitive processes and may improve decision-making in complex choices. The waiting period is followed by a VCM. The VCM should be implemented in a later phase of the PDA because this will allow people to consider all relevant information on the decision, instead of solely basing their decision on the aspects of the decision problem that they are familiar with. People can use the information that was provided in the previous phase of the PDA and deliberate on which elements are important to them and suit their values and preferences. The VCM is placed after the waiting period because this will prevent that choices in the VCM are driven by strong emotions [38].

6. Conclusions

Given the complex characteristics of new cancer screening programs that include epigenetic testing and psychological insight that have shown that people have limited rationality and cognitive biases, new strategies are needed to ensure that informed consent reflects an autonomous choice. This processing of information is necessary because solely supplying patients with information insufficiently promotes a choice that is made with understanding and in line with their values. We proposed a framework to guide the development of patient decision aids within the tiered layered staged informed consent model to support informed consent procedures and promote autonomous choices. This framework can be used as guidance, but cannot guarantee that patients actually use the resulting PDA and make autonomous choices. Hence, the responsibility to verify whether a patient is well-informed and the responsibility to obtain informed consent lies with the health care professional or government, depending on the screening context. In addition, more empirical research is needed on how people respond to risk information during medical decision making, and on future strategies to operationalize deliberative and intuitive processes in PDAs.

References

1. Winawer, S.; Fletcher, R.; Rex, D.; Bond, J.; Burt, R.; Ferrucci, J.; Ganiats, T.; Levin, T.;Woolf, S.; Johnson, D.; et al. Colorectal cancer screening and surveillance: Clinical guidelines and rationale-update based on new evidence. Gastroenterology 2003, 124, 544–560. [CrossRef] [PubMed]
2. Goldie, S.J.; Kim, J.J.; Wright, T.C. Cost-effectiveness of human papillomavirus DNA testing for cervical cancer screening in women aged 30 years or more. Obs. Gynecol. 2004, 103, 619–631. [CrossRef] [PubMed]
3. Fulfilling the Potential of Cancer Prevention and Early Detection; National Academies Press (US):Washington, DC, USA, 2003.
4. Goel, V. Appraising organised screening programmes for testing for genetic susceptibility to cancer. BMJ 2001, 322, 1174–1178. [CrossRef] [PubMed]
5. Pashayan, N.; Reisel, D.; Widschwendter, M. Integration of genetic and epigenetic markers for risk stratification: Opportunities and challenges. Per Med. 2016, 13, 93–95. [CrossRef] [PubMed]
6. Hood, L.; Friend, S.H. Predictive, personalized, preventive, participatory (p4) cancer medicine. Nat. Rev. Clin. Oncol. 2011, 8, 184–187. [CrossRef] [PubMed] J. Pers. Med. 2019, 9, 14 8 of 9
7. Widschwendter, M.; Jones, A.; Evans, I.; Reisel, D.; Dillner, J.; Sundstrom, K.; Steyerberg, E.W.; Vergouwe, Y.; Wegwarth, O.; Rebitschek, F.G.; et al. Epigenome-based cancer risk prediction: Rationale, opportunities and challenges. Nat. Rev. Clin. Oncol. 2018, 15, 292–309. [CrossRef] [PubMed]
8. Stacey, D.; Legare, F.; Lewis, K.; Barry, M.J.; Bennett, C.L.; Eden, K.B.; Holmes-Rovner, M.; Llewellyn-Thomas, H.; Lyddiatt, A.; Thomson, R.; et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst. Rev. 2017, 4, CD001431. [CrossRef] [PubMed]
9. Stacey, D.; Samant, R.; Bennett, C. Decision making in oncology: A review of patient decision aids to support patient participation. CA Cancer J. Clin. 2008, 58, 293–304. [CrossRef] [PubMed]
10. O'Brien, M.A.; Whelan, T.J.; Villasis-Keever, M.; Gafni, A.; Charles, C.; Roberts, R.; Schiff, S.; Cai, W. Are cancer-related decision aids effective? A systematic review and meta-analysis. J. Clin. Oncol. 2009, 27, 974–985. [CrossRef] [PubMed]
11. Neuman, H.B.; Charlson, M.E.; Temple, L.K. Is there a role for decision aids in cancer-related decisions? Crit. Rev. Oncol./Hematol. 2007, 62, 240–250. [CrossRef] [PubMed]
12. Vos, I.M.L.; Schermer, M.H.N.; Bolt, I.L.L.E. Recent insights into decision-making and their implications for informed consent. J. Med. Ethics 2018, 44, 734. [CrossRef] [PubMed]
13. Kanwal, R.; Gupta, S. Epigenetic modifications in cancer. Clin. Genet. 2012, 81, 303–311. [CrossRef] [PubMed]
14. Dupont, C.; Armant, D.R.; Brenner, C.A. Epigenetics: Definition, mechanisms and clinical perspective. Semin Reprod. Med. 2009, 27, 351–357. [CrossRef] [PubMed]
15. Shiovitz, S.; Korde, L.A. Genetics of breast cancer: A topic in evolution. Ann. Oncol. 2015, 26, 1291–1299. [CrossRef] [PubMed]

16. Verma, M.; Khoury, M.J.; Ioannidis, J.P. Opportunities and challenges for selected emerging technologies in cancer epidemiology: Mitochondrial, epigenomic, metabolomic, and telomerase profiling. Cancer Epidemiol. Biomark. Prev. 2013, 22, 189–200. [CrossRef] [PubMed]
17. Forecee Female Caner Prediction Using Cervical Omics to Individualise Screening and Prevention. Available online: http://www.forecee.eu/ (accessed on 20 November 2017).
18. Van El, C.G.; Cornel, M.C.; Borry, P.; Hastings, R.J.; Fellmann, F.; Hodgson, S.V.; Howard, H.C.; Cambon-Thomsen, A.; Knoppers, B.M.; Meijers-Heijboer, H.; et al. Whole-genome sequencing in health care: Recommendations of the european society of human genetics. Eur J. Hum. Genet. 2013, 21, 580–584. [CrossRef] [PubMed]
19. Ubel, P.A. Beyond costs and benefits: Understanding how patients make health care decisions. Oncologist 2010, 15, 5–10. [CrossRef] [PubMed]
20. Marteau, T.M.; Hollands, G.J.; Fletcher, P.C. Changing human behavior to prevent disease: The importance of targeting automatic processes. Science 2012, 337, 1492–1495. [CrossRef] [PubMed]
21. Hall, A.E.; Chowdhury, S.; Hallowell, N.; Pashayan, N.; Dent, T.; Pharoah, P.; Burton, H. Implementing risk-stratified screening for common cancers: A review of potential ethical, legal and social issues. J. Public Health 2014, 36, 285–291. [CrossRef] [PubMed]
22. Kegley, J.A.K. Challenges to informed consent. EMBO Rep. 2004, 5, 832–836. [CrossRef] [PubMed] Faden, R.R.; Beauchamp, T.L.; King, N.M.P. A History and Theory of Informed Consent; Oxford University Press: Oxford, UK, 1986.
23. Beauchamp, T.L. Informed consent: Its history, meaning, and present challenges. Camb. Q. Healthc. Ethics 2011, 20, 515–523. [CrossRef] [PubMed]
24. Dworkin, G. The Theory and Practice of Autonomy; Cambridge University Press: Cambridge, UK, 1988
25. Berger, J. Informed consent is inadequate and shared decision making is ineffective: Arguing for the primacy of authenticity in decision-making paradigms. Am. J. Bioeth. 2017, 17, 45–47. [CrossRef] [PubMed
26. Bradbury, A.R.; Patrick-Miller, L.; Domchek, S. Multiplex genetic testing: Reconsidering utility and informed consent in the era of next-generation sequencing. Genet. Med. 2015, 17, 97–98. [CrossRef] [PubMed]
27. Ormond, K.E.; Iris, M.; Banuvar, S.; Minogue, J.; Annas, G.J.; Elias, S. What do patients prefer: Informed consent models for genetic carrier testing. J. Genet. Couns. 2007, 16, 539–550. [CrossRef] [PubMed]
28. Tversky, A.; Kahneman, D. Judgment under uncertainty: Heuristics and biases. Science 1974, 185, 1124–1131. [CrossRef] [PubMed]
29. Wegwarth, O.; Widschwendter, M.; Cibula, D.; Sundstrom, K.; Portuesi, R.; Lein, I.; Rebitschek, F.G. What do European women know about their female cancer risks and cancer screening? A cross-sectional online intervention survey in five european countries. BMJ Open 2018. [CrossRef] [PubMed] J. Pers. Med. 2019, 9, 14 9 of 9
30. Bunnik, E.M.; Janssens, A.C.J.W.; Schermer, M.H.N. A tiered-layered-staged model for informed consent in personal genome testing. Eur. J. Hum. Genet. 2013, 21, 596–601. [CrossRef] [PubMed]

31. Bunnik, E.M.; Janssens, A.C.; Schermer, M.H. Informed consent in direct-to-consumer personal genome testing: The outline of a model between specific and generic consent. Bioethics 2014, 28, 343–351. [CrossRef] [PubMed]
32. Eiseman, E.; Bloom, G.; Brower, J.; Clancy, N.; Olmsted, S.S. Case Studies of Existing Human Tissue Repositories: "Best Practices" for a Biospecimen Resource for the Genomic and Proteomic Era; RAND Corporation: Santa Monica, CA, USA, 2003.
33. Bijlsma, R.M.; Bredenoord, A.L.; Gadellaa-Hooijdonk, C.G.; Lolkema, M.P.; Sleijfer, S.; Voest, E.E.; Ausems, M.G.; Steeghs, N. Unsolicited findings of next-generation sequencing for tumor analysis within a dutch consortium: Clinical daily practice reconsidered. Eur. J. Hum. Genet. 2016, 24, 1496–1500. [CrossRef][PubMed]
34. Elwyn, G.; O'Connor, A.; Stacey, D.; Volk, R.; Edwards, A.; Coulter, A.; Thomson, R.; Barratt, A.; Barry, M.; Bernstein, S.; et al. Developing a quality criteria framework for patient decision aids: Online international delphi consensus process. BMJ 2006, 333. [CrossRef] [PubMed]
35. Will, I. An introduction to patient decision aids. BMJ 2013, 347. [CrossRef]
36. Fagerlin, A.; Pignone, M.; Abhyankar, P.; Col, N.; Feldman-Stewart, D.; Gavaruzzi, T.; Kryworuchko, J.; Levin, C.A.; Pieterse, A.H.; Reyna, V.; et al. Clarifying values: An updated review. BMC Med. Inform. Decis. Mak. 2013, 13, S8. [CrossRef] [PubMed]
37. De Vries, M.; Fagerlin, A.; Witteman, H.; Scherer, L.D. Combining deliberation and intuition in patientdecision support. Patient Educ. Couns. 2013, 91, 154–160. [CrossRef] [PubMed]
38. Nelson, W.L.; Han, P.K.; Fagerlin, A.; Stefanek, M.; Ubel, P.A. Rethinking the objectives of decision aids: A call for conceptual clarity. Med. Decis. Mak. 2007, 27, 609–618. [CrossRef] [PubMed]
39. Charles, C.; Gafni, A.; Whelan, T.; O'Brien, M.A. Treatment decision aids: Conceptual issues and future directions. Health Expect. 2005, 8, 114–125. [CrossRef] [PubMed]
40. Hunink, M.G.M.;Weinstein, M.C.;Wittenberg, E.; Drummond, M.F.; Pliskin, J.S.;Wong, J.B.; Glasziou, P.P. Decision Making in Health and Medicine: Integrating Evidence and Values, 2nd ed.; Cambridge University Press: Cambridge, UK, 2014
41. Malenka, D.J.; Baron, J.A.; Johansen, S.; Wahrenberger, J.W.; Ross, J.M. The framing effect of relative and absolute risk. J. Gen. Intern. Med. 1993, 8, 543–548. [CrossRef] [PubMed]

Chapter 3

Cost-effectiveness of prophylactic hysterectomy in first-degree female relatives with Lynch syndrome of patients diagnosed with colorectal cancer in the United States: a microsimulation study

Maaike Alblas, Elisabeth F.P. Peterse, Mengmeng Du, Ann G. Zauber, Ewout W. Steyerberg, Nikki van Leeuwen, Iris Lansdorp-Vogelaar

Cancer Medicine. 2021

ABSTRACT

Background: To evaluate the cost-effectiveness of prophylactic hysterectomy (PH) in women with Lynch syndrome (LS)

Methods: We developed a microsimulation model incorporating the natural history for the development of hyperplasia with and without atypia into endometrial cancer (EC) based on the MISCAN-framework. We simulated women identified as first-degree relatives (FDR) with LS of colorectal cancer patients after universal testing for LS. We estimated costs and benefits of offering this cohort PH, accounting for reduced quality of life after PH and for having EC. Three minimum ages (30/35/40) and three maximum ages (70/75/80) were compared to no PH.

Results: In the absence of PH, the estimated number of EC cases was 300 per 1,000 women with LS. Total associated costs for treatment of EC were $5.9 million. Offering PH to FDRs aged 40-80 years was considered optimal. This strategy reduced the number of endometrial cancer cases to 5.4 (-98%), resulting in 516 quality-adjusted life years (QALY) gained and increasing the costs (treatment of endometrial cancer and PH) to $15.0 million (+154%) per 1,000 women. PH from earlier ages was more costly and resulted in fewer QALYs, although this finding was sensitive to disutility for PH.

Conclusions: Offering PH to 40- to 80-year-old women with LS is expected to add 0.5 QALY per person at acceptable costs. Women may decide to have PH at a younger age, depending on their individual disutility for PH and premature menopause.

INTRODUCTION

It has been standard policy for years to try and identify Lynch Syndrome (LS) mutation carriers among colorectal cancer (CRC) patients. Initially, this was done using family history criteria, but since the past decade, universal reflex testing of tumors of CRC patients for mismatch repair deficiency has become increasingly accepted. The aim of this practice is to identify first-degree relatives (FDR) with LS, in order to provide them with preventive interventions.[1-5] LS is a hereditary condition that causes a substantial risk of both colorectal cancer (30%–60%) and endometrial cancer (17%–60%).[6-9] It is estimated that approximately 1 in 300 individuals have LS in the United States (US). [10-12] The practice of universal testing for LS and offering FDR with LS intensive colonoscopy screening for colorectal cancer has shown to be (cost-)effective.13, 14 Yearly endometrial sampling from age 30–35 years onwards might be considered a possible screening strategy for female carriers, but there is no consensus on the effectiveness and impact on quality of life of this strategy.[15] Prophylactic hysterectomy combined with oophorectomy (further referred to as prophylactic hysterectomy, PH) when childbearing is completed has been suggested as a preventive strategy.[4, 16] It might prevent nearly all endometrial cancer cases and deaths in women with LS.[4, 16] However, little is known about its cost-effectiveness and the optimal age range. Determining this optimal age range requires to consider different elements that are associated with PH, such as costs and quality of life. One study using a Markov model showed that offering prophylactic hysterectomy from age 40 is cost-effective, but these results were based on a single-age cohort and only a limited number of strategies (two minimum ages and no maximum age).[17] In reality, the age distribution of identified LS carriers ranges from 11 to 80.[18] This age range is of specific importance because women at higher ages should be able to weigh the benefits and harms of surgery, given that they have not developed symptomatic endometrial cancer. To our knowledge, no previous study has incorporated the age range of LS carriers in their modelling. The aim of this study was to evaluate the cost-effectiveness of offering prophylactic hysterectomy to female FDR with LS, comparing different age ranges to assess optimal age thresholds. Therefore, we developed a microsimulation model for endometrial cancer based on the MISCAN modeling framework.

METHODS

Model specification and assumptions

We used the well-established MISCAN model as a framework to develop the MISCAN Endometrial model. The MISCAN model has been extensively described elsewhere.[19, 20] In short, the MISCAN models simulate a large population of individuals, including life histories from birth to death. The simulations are based on input parameters, which contain both

demographic information and the natural history of the specific disease. The results of the MISCAN models include information on age-specific disease incidence and mortality.

The natural history part of the model is shown in Figure 1 and divides the development of endometrial cancer in three sequential phases: preclinical hyperplasia, preclinical cancer, and clinical cancer.[8] We assumed two types of hyperplasia, of which endometrial hyperplasia without atypia is 6.14 times more frequent than atypical endometrial hyperplasia.[21] The progression of hyperplasia to endometrial cancer differed between hyperplasia without atypia and hyperplasia with atypia, since both have different dwelling times.[21] Dwelling times were derived from Lacey et al. and were estimated with a Weibull distribution.[21] In line with assumptions made for the development of colorectal cancer,[14] preclinical lesions were assumed to progress 10 times faster in LS patients than in the general population. The age-specific onset of endometrial hyperplasia was calibrated to match the incidence of EC for LS women according to Bonadona et al.[8] The survival rates were based on SEER 18 data and were corrected for death due to other causes.[22] Upon of diagnosis of EC, death can occur due to EC or other causes. An elaborative description of our MISCAN model can be found in the Supporting Information Model Appendix.

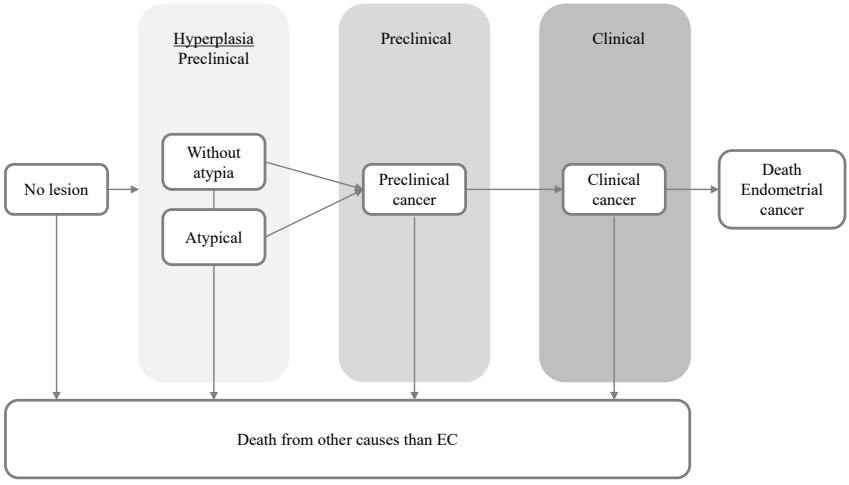

Figure 1 – Natural history model of MISCAN Endometrium model EC Endometrial Cancer

Study population

For each EC prevention strategy, we simulated a population of 10 million Lynch positive women. The target population for prophylactic hysterectomy consisted of FDR with LS of colorectal cancer patients with LS (Figure 2). The age range of the population simulated matched that of FDR with LS in a Dutch study of universal testing of LS in colorectal cancer.18 Individuals were between age 11 and 80 when they were diagnosed with LS. Their median

age was 42 years, with an interquartile range of 31–55 years. In addition, benefits and costs of PH by 5-year age groups were computed.

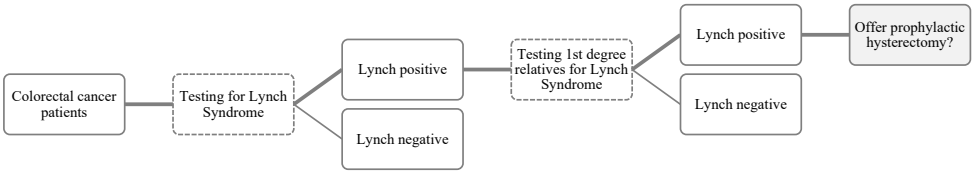

Figure 2 – Flowchart target population for prophylactic hysterectomy

Strategies

Nine different age ranges were modelled with varying ages at which prophylactic hysterectomy was offered as young as 30, 35 or 40 years and as old as age 70, 75, or 80 years. Prophylactic hysterectomy was considered to eliminate the risk of EC completely from date of surgery. We assumed full compliance of every woman who was invited for prophylactic hysterectomy.

Data and assumptions for costs and utilities

An overview of the costs and utilities that were used in the model can be found in Table 1. We assumed that prophylactic hysterectomy reduced the quality of life because of surgically induced menopause. The first month after surgery, quality of life was valued at 0.56, followed by 0.74 in the second and third month after surgery.[23-26] From three months onward, we assumed a utility of 0.88 and corrected the quality of life up to the age of 45, as it is assumed that natural menopause starts at this age which eliminates the negative side effects on quality of life of prophylactic hysterectomy.[17, 24, 27] We also adjusted the quality of life of women diagnosed with EC.[17, 28] The costs of prophylactic hysterectomy are reported as total Medicare reimbursement and include gynecologist fee, anesthesia fee for hysterectomy, pathology fee for uterus, inpatient diagnosis-related group fees, and preoperative lab fees.[29] For the costs of treatment of EC, we assumed 25% of all LS patients receive radiotherapy and 15% of LS patients receive chemotherapy.[16, 30] Furthermore, we included gynecologist fee, anesthesia fee for hysterectomy, pathology fee for uterus, inpatient diagnosis-related group fees, pathology fee for lymph nodes and preoperative lab fees.[29]

Table 1. Model inputs

Variable	Base case	Range	Reference
Cumulative Risk of developing EC before age 80	35%	17-60	Bonadona 2011[8]
Age distribution of FDR[a]	11-80	-	Leenen 2016[18]
Survival probability	Age specific	-	SEER 2009-2013
Ratio of prevalence of hyperplasia without atypia compared to with atypia	6.14	-	Lacey 2010[21]
Life table	Age specific	-	National Vital Statistics Reports 2012[45]
Dwelling time atypical lesions	7.77		Assumption[b]
Dwelling time lesions without atypia	114.40		Assumption[b]
Costs prophylactic hysterectomy[c]	15,276	7,638-30,552	Havrilesky 2009[29]
Costs EC[d]	35,763	17,882-71,526	Schmeler 2006[16] Broaddus 2006[30]
Utility prophylactic hysterectomy	0.88	0.82-0.99	Roberts 2011[23] Bhattacharya 2011[25] Hurskainen 2004[26]
Utility well	1	0.8-1.0	Fryback 1993[46]

EC Endometrial Cancer, FDRs First-degree Relatives [a]The median age was 42 years, with an interquartile range of 31-55 years [b]We derived dwelling times from Lacey et al 2010 with a Weibull distribution. We assumed that for women with Lynch Syndrome, dwelling times were 10 times shorter as for the general population. Values are shown as mean input parameter, dwelling times of lesions that develop into EC will be shorter.[32] [c]Cost reported as total Medicare reimbursement in US dollars. Includes: gynecologist fee, anesthesia fee for hysterectomy, pathology fee for uterus, pathology fee for lymph nodes, inpatient diagnosis-related group fees, preoperative lab fees. [d]We assumed that the costs of treatment for EC are equal to the costs of prophylactic hysterectomy, as treatment of EC usually consists of a hysterectomy.

Outcomes

We determined the effects of offering prophylactic hysterectomy in terms of number of EC deaths, number of prophylactic hysterectomies, life years gained (LYG) and quality-adjusted life years gained (QALYG). We calculated the associated costs for each strategy based on number of prophylactic hysterectomies and total treatment costs for endometrial cancer. We applied a 3% discount rate for both effects and costs to the year in which the women were diagnosed with LS, except for the number of EC cases and deaths. Our analyses were performed with the assumptions described in Table 1. We evaluated average cost-effectiveness ratios (ACERs), which are defined as the difference in costs divided by the difference in QALYG compared to the no prophylactic hysterectomy strategy. Next, the incremental cost-effectiveness ratios (ICERs) of the different strategies were evaluated to determine the optimal strategy. We assumed a willingness-to-pay threshold of 100,000 US dollars per QALY for this analysis.[31, 32]

Sensitivity analyses

To evaluate which assumptions were important drivers for our conclusion, we performed several sensitivity analyses (see range in Table 1). We varied: (1) Quality of life of endometrial cancer, prophylactic hysterectomy and health state well; (2) costs of (prophylactic) hysterectomy; (3) risk of endometrial cancer; and (4) lower life expectancy due to colorectal cancer risk in LS.

RESULTS

In the absence of prophylactic hysterectomy in FDRs with LS, the MISCAN-Endometrium model predicted 300 EC cases and 71 EC deaths per 1,000 women with LS, accounting for the age distribution of the FDR at LS diagnosis. Total associated costs for the treatment of EC were estimated at $5.9 million. Offering these women prophylactic hysterectomy greatly reduced the number of EC cases and deaths, ranging from 0 to 11 and of 0 to 2.9 per 1,000 women, respectively. Although the number of LYG varied relatively little between the different strategies (411–435 per 1,000 women), the number of QALYG was substantially higher for strategies with age 40 as a start age (506–516 per 1,000 women) compared to age 35 (374–384 per 1,000 women) and age 30 (262–272). All strategies with prophylactic hysterectomy were cost-effective compared to no prophylactic hysterectomy, with ACERs

Table 2: Results per 1000 women diagnosed with Lynch syndrome below $50,000 when either LYG or QALYG were used as effectiveness measures (Table 2).

Strategy	EC cases	EC deaths	LYG[a,b]	QALYG[a,b]	Costs[a,] (million US$)	ACER QALYG[a,b]
No prophylactic hysterectomy	300	70.9	-	-	5.9	
30-70	5.6	2.0	426	262	14.1	$31,220
30-75	1.3	0.5	433	269	14.4	$31,618
30-80	0.0	0.0	435	272	14.6	$31,936
35-70	6.6	2.1	423	374	13.7	$20,735
35-75	2.3	2.9	430	381	14.0	$21,228
35-80	1.0	0.2	432	384	14.2	$21,513
40-70	11.0	2.9	411	506	13.2	$14,306
40-75	6.7	1.5	417	514	13.5	$14,768
40-80	5.4	1.0	420	516	13.7	$15,008

Abbreviations: ACER, Average Cost-Effectiveness Ratio; EC, deaths endometrial cancer deaths; LYG, life years gained; QALYG, quality-adjusted life years gained.

a Results are 3% discounted.

b Compared to no prophylactic hysterectomy

When adjusting for quality of life, only strategies in which prophylactic hysterectomy was offered to FDRs after age 40 were efficient strategies; strategies that included prophylactic hysterectomy from age 30 and age 35 were more costly and resulted in fewer quality-adjusted life years gained (Figure 3). The ICERs for ages 40–75 and ages 40–80 were $45,167 and $70,430, respectively. Assuming a willingness-to-pay threshold of $100,000, offering prophylactic hysterectomy to LS women aged 40–80 was considered optimal. Compared to no prophylactic hysterectomy, this strategy would reduce the number of endometrial cancer cases to 5.4 (−98%), resulting in 516 quality-adjusted life years gained and increasing the costs (treatment of endometrial cancer and prophylactic hysterectomy) to $15.0 million (+154%) per 1,000 women. That PH before age 40 is not cost-effective can easily be seen from Table 3. For example, offering PH to women aged 30–34 prevents 77.9 EC deaths compared to 76.2 EC deaths for PH, women aged 40–44 prevents (Table 3), which is an increase of 2.2%. The life-years with PH before age 45 on the other hand increase from approximately 2.5 years to 12.5 years, an increase of 400%. At the other extreme, Table 3 also clearly outlines why PH is still worthwhile even up to age 80: in 75–79 year-olds still more than 40 EC deaths per 1,000 women can be prevented, while the disutility from PH at that age is small, because we only assume disutility in the first three months after surgery.

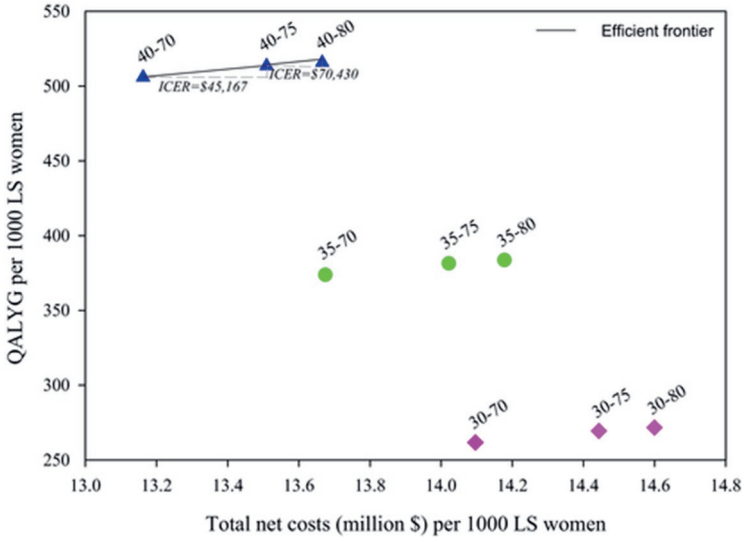

Figure 3 – Efficiency frontier quality-adjusted life years gained QALYG Quality-adjusted life-years gained, LS Lynch syndrome

Sensitivity analyses

The findings of this study were robust for most of our assumptions (Supporting Information Appendix Table 5–14). Only when a higher utility after PH was assumed or life-years gained were considered as the primary outcome, offering prophylactic hysterectomy before age 40

was optimal. However, there were no model-recommended strategies with starting ages below 35 years. The recommended stop age was age 80 in all analyses, except when higher hysterectomy costs were assumed (Table 4).

Table 3: Results per age category (per 1000 women diagnosed with Lynch Syndrome)

Strategy	EC cases prevented	EC deaths prevented	LYG[a, b]	QALYG[a]	Additional Costs[a] (million US$)
30–34	351.8	77.9	460	−489	9.518
35–39	348.5	77.6	510	45	8.811
40–44	339.5	76.2	536	608	8.269
45–49	323.4	73.8	534	918	7.975
50–54	297.1	70.3	502	845	8.087
55–59	258.1	65.8	443	701	8.754
60–64	217.8	60.9	385	558	9.544
65–69	178.8	55.0	320	420	10.372
70–74	142.2	48.0	252	292	11.210

Abbreviations: EC, deaths endometrial cancer deaths; LYG, life years gained; QALYG, quality-adjusted life years gained.
a Results are 3% discounted.

b Earlier PH adds slightly more LYG for women who would otherwise die from EC between this age group and the next. On the other hand, LYG in all women who would be diagnosed with EC after age 35 are discounted for 5 more years and therefore become smaller.

Table 4 – Model-recommended strategies with a willingness-to-pay threshold of $100,000 based on varying input parameters in sensitivity analyses

	Model recommended strategies
Base case	40-80
Base case without adjustment for quality of life	30-80
(Prophylactic) hysterectomy costs	
- 50%	40-80
+ 100%	40-75
Treatment costs endometrial cancer	
−50%	40-80
+100%	40-80
Utility endometrial cancer	
0.68	40-80
Utility prophylactic hysterectomy	
0.82	40-80
0.99	35-80
Risk of endometrial cancer	
17%	40-80
60%	40-80
Accounting for reduced life expectancy due to increased colorectal cancer risk in LS[a]	40-80

Abbreviations: LYG, life years gained; QALYG, quality-adjusted life-years gained.

a MISCAN-Colon was used to generate lifetables that accounted for the increased colorectal cancer mortality of LS women, assuming LS women participated in biennial colonoscopy surveillance from age 25 to age 80.14

DISCUSSION

We evaluated the cost-effectiveness of offering prophylactic hysterectomy to asymptomatic women diagnosed with LS by reflex testing and subsequent cascade testing of FDR with colorectal cancer. Our results show that offering prophylactic hysterectomy to these women is cost-effective at currently accepted standards, and is most cost-effective when offered between age 40 and 80. Depending on an individual disutility for PH and premature menopause, women may decide to undergo PH at a younger age when the perceived impact of PH and premature menopause is small.

Obviously, earlier stop ages were optimal when higher costs of hysterectomy were assumed. The increase in benefits of offering prophylactic hysterectomy to LS women until age 80 rather than age 70 or 75 was relatively small. This may be explained by the median age of diagnosis of endometrial cancer in patients with LS, which is 48 years,[33] while 98% may be diagnosed before the age of 65 years.33 This may support stopping prophylactic hysterectomy before age 70 to prevent potential unnecessary surgery. However, as long as the relative increase in costs is also small, offering prophylactic hysterectomy until age 80 may be considered.

Altering the input parameters for quality of life after PH resulted in the recommendation to start prophylactic hysterectomy at an age younger than 40 years. Women will go into premature menopause as a result of prophylactic hysterectomy, which can result in depression, anxiety, sexual dysfunction and lower self-confidence.[34] We must acknowledge the presence of individual variation in the impact of PH on quality of life during premature menopause. Little is known on this individual variation and specific data on utilities after prophylactic surgery instead of curative surgery is currently lacking. Therefore, empirical data regarding quality of life after prophylactic hysterectomy and the resulting premature menopause are needed to make the quality of life adjustments that are made in our model more robust.

An important strength of this study is that it comprehensively compares the cost-effectiveness of offering prophylactic hysterectomy to women diagnosed with LS for different minimum and maximum ages in a mixed population of different ages. Our results are in line with the results from a prior Markov decision model by Kwon et al,[17] who also showed that offering prophylactic hysterectomy from age 40 was the best strategy. Like us, Kwon et al[17] also showed that the results are highly depended on the inclusion of quality of life in the analyses. In our analyses, starting with prophylactic hysterectomy at age 30 until age 80 prevented all endometrial cancer cases and deaths due to endometrial cancer, leading to a high number of LYG. However, this strategy comes at a high prize in terms of costs and quality of life. Hence, any strategy that starts at the age of 30 or even age 35 was dominated by

strategies that start prophylactic hysterectomy at age 40. In addition, the age when women have their first child is increasing, which might cause women to complete their family at an older age.[35] As a consequence, women may postpone prophylactic hysterectomy. Yang et al[36] identified prophylactic hysterectomy from age 30 as optimal strategy, compared to annual examination. However, no other start ages were tested, which complicates the comparison with the results from our study.

Furthermore, the results of our study are applicable to all asymptomatic women with LS. Although the target population of our study consisted of FDR with LS of colorectal cancer patients with LS, the target population might also be FDR of patients diagnosed with EC. However, the majority of asymptomatic LS patients is identified through a colorectal cancer case in the family, which was therefore the focus of our current analysis. The only parameter in the model that was influenced by this assumption is the age distribution of the asymptomatic LS cases, which was only available for those related to a colorectal cancer patient. As the median ages of colorectal cancer and endometrial cancer diagnoses are comparable, the age distribution of first-degree relatives identified with LS are likely also comparable. Therefore, the results of our study are applicable to all asymptomatic women with LS, regardless of whether they were related to a colorectal cancer or an endometrial cancer patient.

Some limitations of our study should be acknowledged. First, we used the utilities and costs of hysterectomy combined with oophorectomy in our analyses, while we did not incorporate ovarian cancer in our microsimulation model. We have chosen to do so because prophylactic hysterectomy combined with oophorectomy has been recommended as preventive strategy in female patients with LS, given their elevated risk of ovarian cancer (2%–39% life time risk).[16] However, recent studies have shown that ovarian cancer is often detected at an early stage in LS patients, with a relatively good 10-year survival prognosis of 81%.[37-39] Hence, it might be an option to offer younger women the option to undergo a single prophylactic hysterectomy as initial surgery, and to undergo a delayed bilateral salpingo-oophorectomy at menopause. This two-step surgery option might influence the decision of women to undergo prophylactic surgery, since this option does not result in premature menopause. Given the changes in costs and quality of life, some effect on the cost-effectiveness is expected. Based on our sensitivity analysis, in which we assumed a higher utility after prophylactic hysterectomy, we expect that a younger starting age for prophylactic hysterectomy will be the optimal strategy. Future studies are necessary to determine if treatment options such as prophylactic hysterectomy with delayed bilateral salpingo-oophorectomy at menopause are (1) safe for LS patients given their elevated risk of ovarian cancer, and (2) cost-effective.

Second, we assumed that every woman who was invited for prophylactic hysterectomy would undergo this procedure. The model therefore predicted the maximum achievable benefits of prophylactic hysterectomy. Although this implies that the predicted benefits are unrealistic, guidelines should be made based on the benefits that would accrue under perfect rates of adherence to recommendations. Moreover, any change in rates of adherence will have no effects on the ratios that were calculated in our analyses, as the costs and benefits that were used are proportional. Research has shown that FDR of patients with LS underutilize genetic screening, with uptake varying from 15% to 53%.[40] A study on the uptake of bilateral risk-reducing mastectomy and bilateral risk-reducing salpingo-oophorectomy amongst BRCA1/2 mutation carriers showed that uptake was 40% and 45% respectively, and was related to lifetime risk and age.[41] Third, we did not consider other LS-related cancers, such as colorectal or ovarian cancer; due to the lack of data we assumed that apart from an increased EC risk, LS cases have a normal life expectancy. This potentially resulted in an overestimation of life-years gained per EC death prevented. However, our sensitivity analysis showed that our findings were robust when we corrected life expectancy for the increased colorectal cancer mortality in LS. Fourth, the natural history of EC in women with LS is largely unknown. In line with analyses performed for colorectal cancer in LS, we assumed that dwelling times are ten times shorter for women with LS compared to the general population. Fifth, the risk of EC in LS women is uncertain, as estimates vary greatly among studies.[8] We calibrated our model to the largest study that accounted for ascertainment bias,[6-8] and explored higher and lower risk levels in sensitivity analyses. Our results demonstrate that the optimal age range depends on the assumed EC risk for LS cases, which is why future studies are needed to determine the exact risk of EC in LS women. Lastly, we assumed Medicare costs in our analysis while most women might not be Medicare eligible. Also we do not account for non-medical costs such as out-of-pocket costs or time out of work. The current costs might therefore be an underestimation of the costs associated with PH and the treatment of EC. Furthermore, we were unable to find recent cost data to use in our analyses, which might also contribute to an underestimation of the costs as we assumed that the somewhat older cost data were applicable to recent practice. Further studies are necessary to determine these type of costs to enrich existing cost-effectiveness analyses. Nonetheless, sensitivity analysis found our conclusions to be robust for our assumptions on costs and this underestimation will likely not have influenced our conclusions. We did not perform a probabilistic sensitivity analysis since it is not feasible to provide reliable confidence intervals around our estimates due to the lack of data on the distribution of most of the parameters Therefore, we have chosen to conduct several one-way sensitivity analyses. The results of these sensitivity analyses indicate that the findings of our study were robust for most of our assumptions.

Current guidelines in the United States recommend to offer prophylactic hysterectomy to women from age 40 or when childbearing has completed.[3] This is in line with the results from our study and underlines the importance of identifying LS mutation carriers among

colorectal cancer patients and subsequent cascade testing to improve future prospects of these patients in terms of life expectancy and quality of life. However, standards and protocols vary between centers and countries, which may lead to undesired variation.[42] This variation may be caused by conflicting recommendations and protocols on the optimal screening and preventive strategy for LS.[43] Additional information regarding costs and effects of prophylactic hysterectomy, as provided by our study, may aid in the development of uniform protocols and recommendations for the identification of LS mutation carriers. Moreover, our results can inform physicians and women with LS regarding the decision whether or not to perform prophylactic hysterectomy and from which age, which is important in determining the optimal strategy given the preference-sensitive nature of the decisions these patients are facing.

In summary, our study suggests that offering prophylactic hysterectomy to women diagnosed with LS is cost-effective, and is most cost-effective when offered from age 40 until age 80. Individual variation in impact of PH and premature menopause on quality of life must be taken into account and may cause women to start PH earlier. These findings can be used to inform policy makers and clinicians regarding decisions about offering prophylactic hysterectomy to LS women.

REFERENCES

1. Stoffel EM, Mangu PB, Gruber SB, et al. Hereditary colorectal cancer syndromes: American Society of Clinical Oncology Clinical Practice Guideline endorsement of the familial risk-colorectal cancer: European Society for Medical Oncology Clinical Practice Guidelines. J Clin Oncol. 2015;33: 209-217.
2. Rubenstein JH, Enns R, Heidelbaugh J, Barkun A, Clinical Guidelines C. American Gastroenterological Association Institute Guideline on the Diagnosis and Management of Lynch Syndrome. Gastroenterology. 2015;149: 777-782; quiz e716-777.
3. Vasen HF, Blanco I, Aktan-Collan K, et al. Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. Gut. 2013;62: 812-823.
4. Giardiello FM, Allen JI, Axilbund JE, et al. Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US Multi-Society Task Force on colorectal cancer. Gastroenterology. 2014;147: 502-526.
5. Kastrinos F, Ojha RP, Leenen C, et al. Comparison of Prediction Models for Lynch Syndrome Among Individuals With Colorectal Cancer. LID - 10.1093/jnci/djv308 [doi] LID - djv308 [pii].
6. Aarnio M, Sankila R, Pukkala E, et al. Cancer risk in mutation carriers of DNA-mismatch-repair genes. Int J Cancer. 1999;81: 214-218.
7. Dunlop MG, Farrington SM, Carothers AD, et al. Cancer risk associated with germline DNA mismatch repair gene mutations. Hum Mol Genet. 1997;6: 105-110.
8. Bonadona V, Bonaiti B, Olschwang S, et al. Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. JAMA. 2011;305: 2304-2310.
9. Cenin DR, Naber SK, Lansdorp-Vogelaar I, et al. Costs and outcomes of Lynch syndrome screening in the Australian colorectal cancer population. J Gastroenterol Hepatol. 2018;33: 1737-1744.
10. Win AK, Jenkins MA, Dowty JG, et al. Prevalence and Penetrance of Major Genes and Polygenes for Colorectal Cancer. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2017;26: 404-412.
11. Boland CR, Shike M. Report from the Jerusalem workshop on Lynch syndrome-hereditary nonpolyposis colorectal cancer. Gastroenterology. 2010;138: 2197.e2191-2197.e21977.
12. Hampel H, de la Chapelle A. The search for unaffected individuals with Lynch syndrome: do the ends justify the means? Cancer prevention research (Philadelphia, Pa). 2011;4: 1-5.
13. Mvundura M, Grosse SD, Hampel H, Palomaki GE. The cost-effectiveness of genetic testing strategies for Lynch syndrome among newly diagnosed patients with colorectal cancer. Genetics In Medicine. 2010;12: 93.
14. Peterse EFP, Naber SK, Daly C, et al. Cost-effectiveness of Active Identification and Subsequent Colonoscopy Surveillance of Lynch Syndrome Cases. Clin Gastroenterol Hepatol. 2019.
15. Lindor NM, Petersen GM, Hadley DW, et al. Recommendations for the Care of Individuals With an Inherited Predisposition to Lynch SyndromeA Systematic Review. JAMA. 2006;296: 1507-1517.
16. Schmeler KM, Lynch HT, Chen LM, et al. Prophylactic surgery to reduce the risk of gynecologic cancers in the Lynch syndrome. N Engl J Med. 2006;354: 261-269.

17. Kwon JS, Sun Cc Fau - Peterson SK, Peterson Sk Fau - White KG, et al. Cost-effectiveness analysis of prevention strategies for gynecologic cancers in Lynch syndrome.
18. Leenen CH, Goverde A, de Bekker-Grob EW, et al. Cost-effectiveness of routine screening for Lynch syndrome in colorectal cancer patients up to 70 years of age. Genet Med. 2016;18: 966-973.
19. van Hees F, Habbema JD, Meester RG, Lansdorp-Vogelaar I, van Ballegooijen M, Zauber AG. Should colorectal cancer screening be considered in elderly persons without previous screening? A cost-effectiveness analysis. Ann Intern Med. 2014;160: 750-759.
20. Lansdorp-Vogelaar I, van Ballegooijen M, Zauber AG, Habbema JD, Kuipers EJ. Effect of rising chemotherapy costs on the cost savings of colorectal cancer screening. J Natl Cancer Inst. 2009;101: 1412-1422.
21. Lacey JV, Jr., Sherman ME, Rush BB, et al. Absolute risk of endometrial carcinoma during 20-year follow-up among women with endometrial hyperplasia. J Clin Oncol. 2010;28: 788-792.
22. Cenin DR, St John J, Slevin T, Ledger MJ, Lansdorp-Vogelaar I. Optimising the expansion of the national bowel cancer screening program. The Medical journal of Australia. 2014;201: 456.
23. Roberts TE, Tsourapas A, Middleton LJ, et al. Hysterectomy, endometrial ablation, and levonorgestrel releasing intrauterine system (Mirena) for treatment of heavy menstrual bleeding: cost effectiveness analysis. BMJ (Clinical research ed). 2011;342: d2202-d2202.
24. Miller JD, Lenhart GM, Bonafede MM, Lukes AS, Laughlin-Tommaso SK. Cost-Effectiveness of Global Endometrial Ablation vs. Hysterectomy for Treatment of Abnormal Uterine Bleeding: US Commercial and Medicaid Payer Perspectives. Population health management. 2015;18: 373-382.
25. Bhattacharya S, Middleton LJ, Tsourapas A, et al. Hysterectomy, endometrial ablation and Mirena(R) for heavy menstrual bleeding: a systematic review of clinical effectiveness and cost-effectiveness analysis. Health Technol Assess. 2011;15: iii-xvi, 1-252.
26. Hurskainen R, Teperi J, Rissanen P, et al. Clinical outcomes and costs with the levonorgestrel-releasing intrauterine system or hysterectomy for treatment of menorrhagia: randomized trial 5-year follow-up. JAMA. 2004;291: 1456-1463.
27. Fryback DG, Dasbach Ej Fau - Klein R, Klein R Fau - Klein BE, et al. The Beaver Dam Health Outcomes Study: initial catalog of health-state quality factors.
28. Sun CC, Peterson SK, White KG, et al. Preferences for cancer prevention strategies (CPS) in women with hereditary nonpolyposis colectal cancer (HNPCC). Journal of Clinical Oncology. 2006;24: 1018-1018.
29. Havrilesky LJ, Maxwell GL, Myers ER. Cost-effectiveness analysis of annual screening strategies for endometrial cancer. Am J Obstet Gynecol. 2009;200: 640 e641-648.
30. Broaddus RR, Lynch Ht Fau - Chen L-M, Chen Lm Fau - Daniels MS, et al. Pathologic features of endometrial carcinoma associated with HNPCC: a comparison with sporadic endometrial carcinoma.
31. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. JAMA. 2016;316: 1093-1103.

32. Neumann PJ, Cohen JT, Weinstein MC. Updating cost-effectiveness--the curious resilience of the $50,000-per-QALY threshold. N Engl J Med. 2014;371: 796-797.
33. Vasen HF, Watson P, Mecklin JP, et al. The epidemiology of endometrial cancer in hereditary nonpolyposis colorectal cancer. Anticancer Res. 1994;14: 1675-1678.
34. Deeks AA, Gibson-Helm M, Teede H, Vincent A. Premature menopause: a comprehensive understanding of psychosocial aspects. Climacteric. 2011;14: 565-572.
35. Mathews TJ, Hamilton BE. Delayed childbearing; more women are having their first child later in life2009.
36. Yang KY, Caughey AB, Little SE, Cheung MK, Chen L-M. A cost-effectiveness analysis of prophylactic surgery versus gynecologic surveillance for women from hereditary non-polyposis colorectal cancer (HNPCC) Families. Familial Cancer. 2011;10: 535-543.
37. Ketabi Z, Bartuma K Fau - Bernstein I, Bernstein I Fau - Malander S, et al. Ovarian cancer linked to Lynch syndrome typically presents as early-onset, non-serous epithelial tumors.
38. Crijnen TE, Janssen-Heijnen Ml Fau - Gelderblom H, Gelderblom H Fau - Morreau J, et al. Survival of patients with ovarian cancer due to a mismatch repair defect.
39. Grindedal EM, Renkonen-Sinisalo L, Vasen H, et al. Survival in women with MMR mutations and ovarian cancer: a multicentre study in Lynch syndrome kindreds. Journal of Medical Genetics. 2010;47: 99-102.
40. Sharaf RN, Myer P, Stave CD, Diamond LC, Ladabaum U. Uptake of Genetic Testing by Relatives of Lynch Syndrome Probands: A Systematic Review. Clinical Gastroenterology and Hepatology. 2013;11: 1093-1100.
41. Evans DGR, Lalloo F, Ashcroft L, et al. Uptake of Risk-Reducing Surgery in Unaffected Women at High Risk of Breast and Ovarian Cancer Is Risk, Age, and Time Dependent. Cancer Epidemiology Biomarkers & Prevention. 2009;18: 2318-2324.
42. Jarvinen HJ, Aarnio M, Mustonen H, et al. Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. Gastroenterology. 2000;118: 829-834.
43. Daly C, Rotenberg C, Facey M, Baker NA, Baxter NN. Reflex Lynch syndrome screening by example: A review of existing programs. Journal of Clinical Oncology. 2015;33: 543-543.
44. Rahm AK, Cragun D, Hunter JE, et al. Implementing universal Lynch syndrome screening (IMPULSS): protocol for a multi-site study to identify strategies to implement, adapt, and sustain genomic medicine programs in different organizational contexts. BMC Health Serv Res. 2018;18: 824.
45. Martin JA, Hamilton BE, Osterman MJ, Curtin SC, Matthews TJ. Births: final data for 2012. Natl Vital Stat Rep. 2013;62: 1-68.
46. Fryback DG, Dasbach EJ, Klein R, et al. The Beaver Dam Health Outcomes Study: initial catalog of health-state quality factors. Med Decis Making. 1993;13: 89-102.
47. van Ballegooijen M, Rutter CM, Knudsen AB, et al. Clarifying differences in natural history between models of screening: the case of colorectal cancer. Med Decis Making. 2011;31: 540-549.

Appendix 1

Model Appendix

Microsimulation model structure

The Microsimulation Screening Analysis (MISCAN) program was first developed in 1985 to evaluate the effects of screening on disease.1 Since then, the MISCAN program has been used to quantify the effects of primary and secondary prevention for cancers of the breast, colon, cervix, esophagus, pancreas, prostate, and lung.[2-8]

MISCAN Endometrial is a stochastic, semi-Markov, microsimulation model for endometrial cancer (EC) programmed in Delphi (Borland Software Corporation, Scotts Valley, California, United States). It can be used to explain and predict trends in EC incidence and mortality and to quantify the effects and costs of primary prevention of EC, screening for EC, and prophylactic hysterectomy. The term 'microsimulation' implies that the individuals are moved through the model one at a time, rather than as proportions of a cohort. The term 'semi-Markov' implies that MISCAN Endometrial, unlike traditional Markov models, does not assume annual state transitions; instead it generates durations in states, allowing future state transitions to depend on past transitions, and thereby increases model flexibility and computational performance. The term 'stochastic' implies that the model determines the states and corresponding durations by drawing from probability distributions, rather than using fixed values. Hence, the results of the model are subject to random variation. The version of the MISCAN Endometrial model used for this manuscript consists of a demography module, a natural history module and a prophylactic hysterectomy module.

Demography module

Using birth- and life-tables, MISCAN Endometrial draws a date of birth and a date of non-EC death for each woman simulated. Birth tables were based on Leenen et al.,[9] reflecting the age range of first-degree relatives of individuals diagnosed with lynch syndrome by universal testing of lynch syndrome in colorectal cancer. Life tables were based on National Vital Statistics Reports 2012,[10] reflecting the life expectancy of women in the US. The maximum age an individual can achieve is assumed to be 100 years.

Natural history module

Transitions

All women are born without lesions. As each simulated woman ages, a woman may develop endometrial hyperplasia, either atypical or without atypia (Figure 1). Hyperplasia without

atypia was assumed to be 6.14 times more frequent than atypical hyperplasia.[11] Hyperplasia may or may not progress into preclinical cancer, i.e cancer not yet giving symptoms. Over time, the preclinical cancer may start to give symptoms, resulting in cancer diagnosis. Upon diagnoses, women move to the clinical cancer state. After clinical diagnosis, EC survival is simulated using age-specific survival estimates based on 2009-2013 data from SEER 18.[12] As it was assumed that women diagnosed with EC will receive hysterectomy combined with oophorectomy as part of their treatment, women can only develop endometrial hyperplasia and EC once during their lifetime. The date of death for individuals with EC is set to the earliest simulated death due either to EC or another cause ('Demography module').

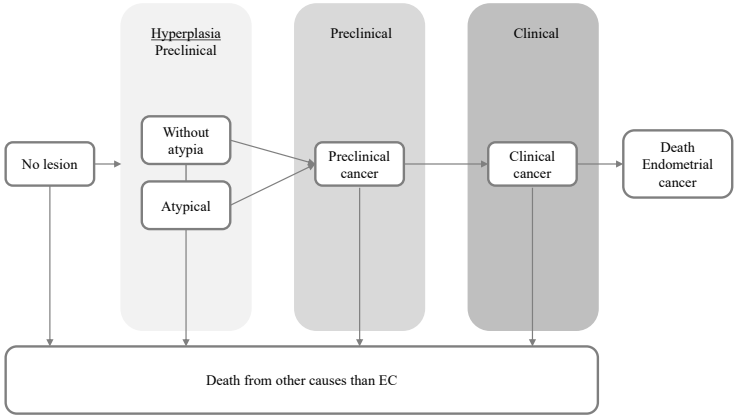

Figure 1: The stages of disease in the semi-Markov model.

Transition rates and durations

A woman's risk of developing hyperplasia depends on the woman's age and a personal Gamma-distributed risk index (non-homogeneous Poisson process). The age-specific onset of hyperplasia was calibrated to the study by Bonadona et al,[13] assuming a 35% risk of women diagnosed with lynch syndrome to develop EC before the age of 80. The progression of hyperplasia to endometrial cancer differed between hyperplasia without atypia and hyperplasia with atypia, since both have different dwelling times.[11] Dwelling times were derived from Lacey et al. and were estimated with a Weibull distribution.[11] In line with assumptions made for the development of colorectal cancer,[14] preclinical lesions were assumed to progress 10 times faster in lynch syndrome patients than in the general population. Mean dwelling times for atypical hyperplasia and hyperplasia without atypia were 7.77 years and 114 years, respectively, reflecting that not all endometrial hyperplasia progresses to cancer.

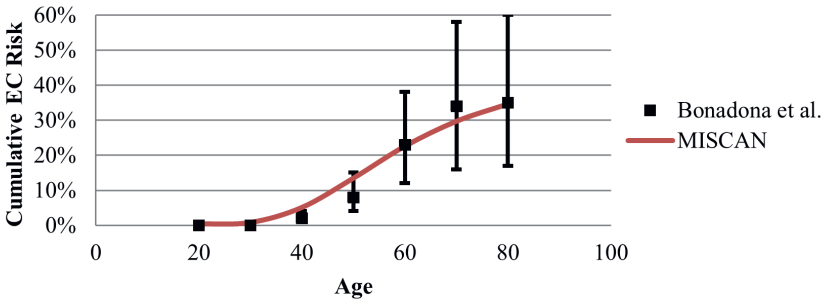

Figure 2: Simulated versus observed endometrial cancer incidence, based on the study by Bonadona et al.13

Prophylactic hysterectomy module

In the prophylactic hysterectomy module, women of a certain age (e.g., 40-80 years) undergo prophylactic hysterectomy. Women who underwent a prophylactic hysterectomy are no longer at risk for developing endometrial hyperplasia and EC. If any preclinical hyperplasia was present at the time of the prophylactic hysterectomy, it is assumed that this is removed and that this no longer impacts a woman's life expectancy.

Integrating modules

In Figure 3, the life history of an example patient is shown; this figure can be used to demonstrate how the different modules are integrated and how the benefit of prophylactic hysterectomy is quantified. For each individual simulated, the demography module first generates a date of birth and a date of non-EC death, creating a life-history without endometrial hyperplasia or EC. Then, the natural history module comes into play, generating onset of disease. For some women, the onset of disease falls after their age of death of other causes, and the woman will not develop any disease. In the example in Figure 3, the simulated woman develops atypical hyperplasia at a relatively young age. This hyperplasia progresses into preclinical cancer, which is diagnosed because of symptoms and results in EC death before non-EC death would have occurred. In the prophylactic hysterectomy module, a prophylactic hysterectomy is simulated, indicated by the blue arrow. In this example, the prophylactic hysterectomy is performed when the woman already developed atypical hyperplasia. As a consequence of the prophylactic hysterectomy, both EC and EC death are prevented. Hence, integrating all three modules, prophylactic hysterectomy prolongs life by the amount indicated by the green arrow.

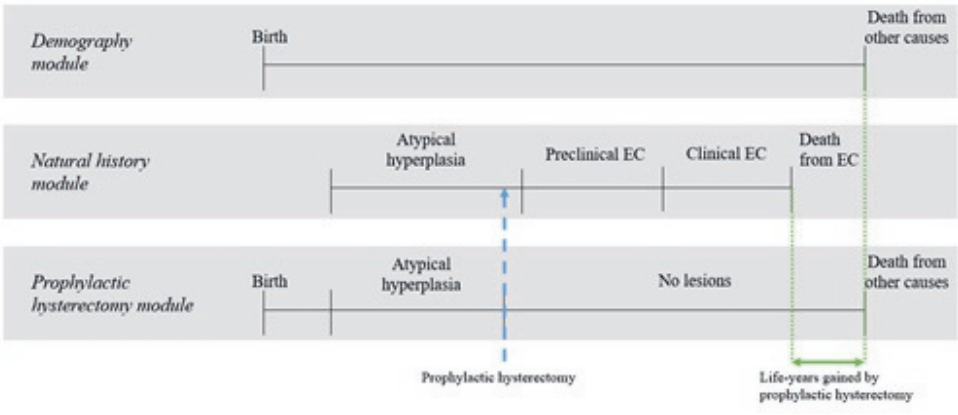

Figure 3: Integrating MISCAN modules for one example patient.

References

1. Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed. 1985;20(1):79-93.
2. Draisma G, Etzioni R, Tsodikov A, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. J Natl Cancer Inst. 2009;101(6):374-383.
3. Loeve F, Boer R, van Oortmarssen GJ, van Ballegooijen M, Habbema JD. The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening. Comput Biomed Res. 1999;32(1):13-33.
4. Ten Haaf K, de Koning HJ. Should Never-Smokers at Increased Risk for Lung Cancer Be Screened? J Thorac Oncol. 2015;10(9):1285-1291.
5. van den Akker-van Marle ME, van Ballegooijen M, van Oortmarssen GJ, Boer R, Habbema JD. Cost-effectiveness of cervical cancer screening: comparison of screening policies. J Natl Cancer Inst. 2002;94(3):193-204.
6. van der Maas PJ, de Koning HJ, van Ineveld BM, et al. The cost-effectiveness of breast cancer screening. Int J Cancer. 1989;43(6):1055-1060.
7. Kroep S, Lansdorp-Vogelaar I, Rubenstein JH, et al. An Accurate Cancer Incidence in Barrett's Esophagus: A Best Estimate Using Published Data and Modeling. Gastroenterology. 2015;149(3):577-585 e574; quiz e514-575.
8. Koopmann BDM, Harinck F, Kroep S, et al. Identifying key factors for the effectiveness of pancreatic cancer screening: A model-based analysis. Int J Cancer. 2021.
9. Leenen CH, Goverde A, de Bekker-Grob EW, et al. Cost-effectiveness of routine screening for Lynch syndrome in colorectal cancer patients up to 70 years of age. Genet Med. 2016;18(10):966-973.
10. Martin JA, Hamilton BE, Osterman MJ, Curtin SC, Matthews TJ. Births: final data for 2012. Natl Vital Stat Rep. 2013;62(9):1-68.
11. Lacey JV, Jr., Sherman ME, Rush BB, et al. Absolute risk of endometrial carcinoma during 20-year follow-up among women with endometrial hyperplasia. J Clin Oncol. 2010;28(5):788-792.
12. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2015 Sub (1973-2013 varying) - Linked To County Attributes - Total U.S., 1969-2014 Counties.
13. Bonadona V, Bonaiti B, Olschwang S, et al. Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. JAMA. 2011;305(22):2304-2310.
14. Peterse EFP, Naber SK, Daly C, et al. Cost-effectiveness of Active Identification and Subsequent Colonoscopy Surveillance of Lynch Syndrome Cases. Clin Gastroenterol Hepatol. 2019.

Appendix 2

Table 5 : Results sensitivity analysis: -50% hysterectomy costs (per 1000 women diagnosed with Lynch syndrome)

Strategy	EC cases	EC deaths	LYG[a]	QALYG[a]	Costs[a,] (million US$)	ACER QALYG[a,b]	ICER QALYG
No prophylactic hysterectomy	300	70.9	-	-	5.9		
40-70	11.0	2.9	411	506	6.7	$1,600	$14,306
40-75	6.7	1.5	417	514	6.8	$1,796	$45,232
40-80	5.4	1.0	420	516	6.9	$1,901	$70,452
35-70	6.6	2.1	423	374	6.9	$2,694	Dominated
35-75	2.3	2.9	430	381	7.0	$2,935	Dominated
30-70	5.6	2.0	426	262	7.1	$4,600	Dominated
35-80	1.0	0.2	432	384	7.1	$3,070	Dominated
30-75	1.3	0.5	433	269	7.2	$4,887	Dominated
30-80	0.0	0.0	435	272	7.3	$5,061	Dominated

EC deaths endometrial cancer deaths, LYG life years gained, QALYG quality-adjusted life years gained, ACER Average Cost-Effectiveness Ratio

a Results are 3% discounted

b Compared to no prophylactic hysterectomy

Table 6 : Results sensitivity analysis: +100% hysterectomy costs (per 1000 women diagnosed with Lynch syndrome)

Strategy	EC cases	EC deaths	LYG[a]	QALYG[a]	Costs[a,] (million US$)	ACER QALYG[a,b]	ICER QALYG
No prophylactic hysterectomy	300	70.9	-	-	5.9		
40-70	11.0	2.9	411	506	26.0	$39,718	$39,718
40-75	6.7	1.5	417	514	26.8	$40,714	$106,335
35-70	6.6	2.1	423	374	27.2	$56,817	Dominated
40-80	5.4	1.0	420	516	27.2	$41,223	$159,019
35-75	2.3	2.9	430	381	28.0	$57,813	Dominated
30-70	5.6	2.0	426	262	28.0	$84,461	Dominated
35-80	1.0	0.2	432	384	28.3	$58,399	Dominated
30-75	1.3	0.5	433	269	28.8	$85,081	Dominated
30-80	0.0	0.0	435	272	29.2	$85,685	Dominated

EC deaths endometrial cancer deaths, LYG life years gained, QALYG quality-adjusted life years gained, ACER Average Cost-Effectiveness Ratio

a Results are 3% discounted

b Compared to no prophylactic hysterectomy

Table 7 : Results sensitivity analysis: utility endometrial cancer 0.68 (per 1000 women diagnosed with Lynch syndrome)

Strategy	EC cases	EC deaths	LYG[a]	QALYG[a]	Costs[a,] (million US$)	ACER QALYG[a,b]	ICER QALYG
No prophylactic hysterectomy	300	70.9	-	-	5.9		
40-70	11.0	2.9	411	530	13.2	$8,988	$8,988
40-75	6.7	1.5	417	541	13.5	$9,296	$32,470
40-80	5.4	1.0	420	375	13.7	$9,453	$52,237
35-70	6.6	2.1	423	544	13.7	$11,348	Dominated
35-75	2.3	2.9	430	385	14.0	$11,673	Dominated
30-70	5.6	2.0	426	245	14.1	$14,251	Dominated
35-80	1.0	0.2	432	388	14.2	$11,847	Dominated
30-75	1.3	0.5	433	255	14.4	$14,584	Dominated
30-80	0.0	0.0	435	258	14.6	$14,776	Dominated

EC deaths endometrial cancer deaths, LYG life years gained, QALYG quality-adjusted life years gained, ACER Average Cost-Effectiveness Ratio

a Results are 3% discounted

b Compared to no prophylactic hysterectomy

Table 8 : Results sensitivity analysis: utility prophylactic hysterectomy 0.82 (per 1000 women diagnosed with Lynch syndrome)

Strategy	EC cases	EC deaths	LYG[a]	QALYG[a]	Costs[a,] (million US$)	ACER QALYG[a,b]	ICER QALYG
No prophylactic hysterectomy	300	70.9	-	-	5.9		
40-70	11.0	2.9	411	418	13.2	$17,333	$17,333
40-75	6.7	1.5	417	425	13.5	$17,836	$45,211
40-80	5.4	1.0	420	427	13.7	$18,109	$70,444
35-70	6.6	2.1	423	209	13.7	$18,312	Dominated
35-75	2.3	2.9	430	217	14.0	$18,835	Dominated
30-70	5.6	2.0	426	39	14.1	$206,977	Dominated
35-80	1.0	0.2	432	219	14.2	$37,692	Dominated
30-75	1.3	0.5	433	47	14.4	$180,570	Dominated
30-80	0.0	0.0	435	49	14.6	$175,617	Dominated

EC deaths endometrial cancer deaths, LYG life years gained, QALYG quality-adjusted life years gained, ACER Average Cost-Effectiveness Ratio

a Results are 3% discounted

b Compared to no prophylactic hysterectomy

Table 9 : Results sensitivity analysis: utility prophylactic hysterectomy 0.99 (per 1000 women diagnosed with Lynch syndrome)

Strategy	EC cases	EC deaths	LYG[a]	QALYG[a]	Costs[a,] (million US$)	ACER QALYG[a,b]	ICER QALYG
No prophylactic hysterectomy	300	70.9	-	-	5.9		
40-70	11.0	2.9	411	668	13.2	$10,837	$10,837
40-75	6.7	1.5	417	675	13.5	$11,228	$45,727
40-80	5.4	1.0	420	678	13.7	$11,422	$70,467
35-70	6.6	2.1	423	676	13.7	$11,470	Dominated
35-75	2.3	2.9	430	683	14.0	$11,849	$63,304
30-70	5.6	2.0	426	669	14.1	$12,211	Dominated
35-80	1.0	0.2	432	686	14.2	$12,039	$70,467
30-75	1.3	0.5	433	677	14.4	$12,585	Dominated
30-80	0.0	0.0	435	679	14.6	$12,775	Dominated

EC deaths endometrial cancer deaths, LYG life years gained, QALYG quality-adjusted life years gained, ACER Average Cost-Effectiveness Ratio

a Results are 3% discounted

b Compared to no prophylactic hysterectomy

Table 10: Results sensitivity analysis: risk endometrial cancer 17% (per 1000 women diagnosed with Lynch syndrome)

Strategy	EC cases	EC deaths	LYG[a]	QALYG[a]	Costs[a,] (million US$)	ACER QALYG[a,b]	ICER QALYG
No prophylactic hysterectomy	284.0	69.0	-	-	5.4		
40-70	6.3	2.2	363	416	13.1	$18,569	$18,569
40-75	1.8	0.6	370	425	13.5	$19,036	$42,351
40-80	0.4	0.1	372	427	13.6	$19,308	$70,927
35-70	6.0	2.1	363	269	13.7	$30,833	Dominated
35-75	1.5	2.2	371	277	14.0	$31,179	Dominated
30-70	5.9	2.1	364	157	14.1	$55,556	Dominated
35-80	0.1	0.0	373	279	14.2	$31,497	Dominated
30-75	1.4	0.5	371	165	14.5	$54,885	Dominated
30-80	0.0	0.0	373	167	14.6	$55,099	Dominated

EC deaths endometrial cancer deaths, LYG life years gained, QALYG quality-adjusted life years gained, ACER Average Cost-Effectiveness Ratio

a Results are 3% discounted

b Compared to no prophylactic hysterectomy

Table 11: Results sensitivity analysis: risk endometrial cancer 60% (per 1000 women diagnosed with Lynch syndrome)

Strategy	EC cases	EC deaths	LYG[a]	QALYG[a]	Costs[a,] (million US$)	ACER QALYG[a,b]	ICER QALYG
No prophylactic hysterectomy	609.7	136.3	-	-	12.3		
40-70	16.7	4.5	863	1344	13.4	$787	$787
40-75	9.9	2.2	875	1360	13.5	$856	$6,681
40-80	7.9	1.4	878	1363	13.6	$892	$13,572
35-70	10.2	3.3	882	1208	13.9	$1,316	Dominated
35-75	3.5	4.5	893	1224	14.0	$1,386	Dominated
35-80	1.5	0.2	897	1228	14.1	$1,425	Dominated
30-70	8.8	3.1	886	1087	14.4	1,880	Dominated
30-75	2.0	0.8	897	1103	14.5	$1,949	Dominated
30-80	0.0	0.0	900	1107	14.5	$1,990	Dominated

EC deaths endometrial cancer deaths, LYG life years gained, QALYG quality-adjusted life years gained, ACER Average Cost-Effectiveness Ratio

a Results are 3% discounted

b Compared to no prophylactic hysterectomy

Table 12 : Results sensitivity analysis: Accounting for reduced life expectancy due to increased colorectal cancer risk in LS (per 1000 women diagnosed with Lynch syndrome)

Strategy	EC cases	EC deaths	LYG[a]	QALYG[a]	Costs[a,] (million US$)	ACER QALYG[a,b]	ICER QALYG
No prophylactic hysterectomy	297.5	69.9	-	-	5.9		
40-70	10.9	2.9	404	495	13.2	$14,670	$14,670
40-75	6.7	1.5	411	503	13.5	$15,132	$45,244
40-80	5.4	1.0	413	505	13.7	$15,372	$70,569
35-70	6.5	2.1	417	363	13.7	$21,445	Dominated
35-75	2.3	2.9	424	371	14.0	$21,933	Dominated
30-70	5.5	1.9	420	251	14.1	$32,748	Dominated
35-80	1.0	0.2	426	373	14.2	$22,218	Dominated
30-75	1.3	0.5	426	258	14.4	$33,115	Dominated
30-80	0.0	0.0	429	260	14.6	$33,429	Dominated

EC deaths endometrial cancer deaths, LYG life years gained, QALYG quality-adjusted life years gained, ACER Average Cost-Effectiveness Ratio

a Results are 3% discounted

b Compared to no prophylactic hysterectomy

Table 13 : Results sensitivity analysis: -50% costs of treatment EC (per 1000 women diagnosed with Lynch syndrome)

Strategy	EC cases	EC deaths	LYG[a]	QALYG[a]	Costs[a,] (million US$)	ACER QALYG[a,b]	ICER QALYG
No prophylactic hysterectomy	300	70.9	-	-	3.0		
40-70	11.0	2.9	411	506	13.0	$19,859	$19,859
40-75	6.7	1.5	417	514	13.4	$20,357	$53,167
35-70	6.6	2.1	423	374	13.6	$28,409	Dominated
40-80	5.4	1.0	420	516	13.6	$20,611	$79,509
35-75	2.3	2.9	430	381	14.0	$28,906	Dominated
30-70	5.6	2.0	426	262	14.0	$42,230	Dominated
35-80	1.0	0.2	432	384	14.2	$29,199	Dominated
30-75	1.3	0.5	433	269	14.4	$42,540	Dominated
30-80	0.0	0.0	435	272	14.6	$42,842	Dominated

EC deaths endometrial cancer deaths, LYG life years gained, QALYG quality-adjusted life years gained, ACER Average Cost-Effectiveness Ratio

a Results are 3% discounted

b Compared to no prophylactic hysterectomy

Table 14 : Results sensitivity analysis: +100% costs of treatment EC (per 1000 women diagnosed with Lynch syndrome)

Strategy	EC cases	EC deaths	LYG[a]	QALYG[a]	Costs[a,] (million US$)	ACER QALYG[a,b]	ICER QALYG
No prophylactic hysterectomy	300	70.9	-	-	11.9		
40-70	11.0	2.9	411	506	13.5	$3,200	$3,200
40-75	6.7	1.5	417	514	13.7	$3,591	$29,362
40-80	5.4	1.0	420	516	13.8	$3,801	$52,337
35-70	6.6	2.1	423	374	13.9	$5,388	Dominated
35-75	2.3	2.9	430	381	14.1	$5,870	Dominated
30-70	5.6	2.0	426	262	14.3	$6,139	Dominated
35-80	1.0	0.2	432	384	14.3	$9,200	Dominated
30-75	1.3	0.5	433	269	14.5	$9,774	Dominated
30-80	0.0	0.0	435	272	14.6	$10,122	Dominated

EC deaths endometrial cancer deaths, LYG life years gained, QALYG quality-adjusted life years gained, ACER Average Cost-Effectiveness Ratio

a Results are 3% discounted

b Compared to no prophylactic hysterectomy

Part III

Prediction

Chapter 4

Prediction models for endometrial cancer for the general population or symptomatic women: A systematic review

Maaike Alblas, Kimberley B. Velt, Nora Pashayan, Martin Widschwendter, Ewout W. Steyerberg, Yvonne Vergouwe

Critical Reviews in Oncology/Hematology. 2018

ABSTRACT

Objective: To provide an overview of prediction models for the risk of developing endometrial cancer in women of the general population or for the presence of endometrial cancer in symptomatic women.

Methods: We systematically searched the Embase and Pubmed database until September 2017 for relevant publications. We included studies describing the development, the external validation, or the updating of a multivariable model for predicting endometrial cancer in the general population or symptomatic women.

Results: Out of 2756 references screened, 14 studies were included. We found two prediction models for developing endometrial cancer in the general population (risk models) and one extension. Eight studies described the development of models for symptomatic women (diagnostic models), one comparison of the performance of two diagnostic models and two external validation. Sample size varied from 60 (10 with cancer) to 201,811 (855with cancer) women. The age of the women was included as a predictor in almost all models. The risk models included epidemiological variables related to the reproductive history of women, hormone use, BMI, and smoking history. The diagnostic models also included clinical predictors, such as endometrial thickness and recurrent bleeding. The concordance statistic (c), assessing the discriminative ability, varied from 0.68 to 0.77 in the risk models and from 0.73 to 0.957 in the diagnostic models. Methodological information was often limited, especially on the handling of missing data, and the selection of predictors. One risk model and four diagnostic models were externally validated.

Conclusions: Only a few models have been developed to predict endometrial cancer in asymptomatic or symptomatic women. The usefulness of most models is unclear considering methodological shortcomings and lack of external validation. Future research should focus on external validation and extension with new predictors or biomarkers, such as genetic and epigenetic markers.

INTRODUCTION

Endometrial cancer is the sixth most common type of cancer in women worldwide and its incidence has been increasing since 1990 (Ferlay et al., 2013). This increase might be related to improvements in detection in the general population and in diagnostics in women with (postmenopausal) bleeding. Further, in many populations the body mass index (BMI) is rising and several studies have shown that adiposity is the strongest risk factor of endometrial cancer (Kyrgiou et al., 2017; Dixon, 2010; Collaboration NCDRF, 2016; Ng et al., 2014). Other risk factors that are associated with endometrial cancer are higher age, hypertension, diabetes, nulliparity, early menarche, late menopause, oestrogen uptake, and genomic alterations (MacMahon, 1974; Hecht and Mutter, 2006). Combining these risk factors in multivariable prediction models may help to identify women in the general population at high risk of developing endometrial cancer. Prediction models can also facilitate early diagnosis in symptomatic women. Several risk and diagnostics models for endometrial cancer have been developed (Pfeiffer et al., 2013; Wong et al., 2016; Husing et al., 2016; Burbos et al., 2010; Giannella et al., 2014). The models can be used for risk prediction for prevention purposes. Particularly models with modifiable risk factors, such as BMI, hypertension, and oestrogen uptake may facilitate tailored preventive interventions on diet, lifestyle or drug use. This might reduce the incidence of endometrial cancer.

Once endometrial cancer has developed, diagnostic models can be used for early diagnosis. Postmenopausal bleeding and increasing endometrial thickness are the most common symptoms of endometrial cancer and are often considered in these diagnostic models (Gull et al., 2003). The diagnostic models facilitate early diagnosis, which may result in efficient use of diagnostic resources and improved survival. Since no overview of these models has been published so far, we aimed to systematically review multivariable models predicting the risk of endometrial cancer in the general population. We also systematically reviewed models for the presence of endometrial cancer in symptomatic women. We describe the model development, the included predictors, the predicted outcome, and any attempts to external validation to assess the quality of the models and determine if these models are ready for use in practice.

METHODS

Search strategy

The search strategy that was used in this review was based on previous published searches (Damen et al., 2016; Ingui and Rogers, 2001) and other systematic reviews of prediction models (Smit et al., 2015; Meads et al., 2012; Mushkudiani et al., 2008). Specific terms for endometrial cancer were added to the search strategy. The index terms of papers that were

considered relevant were manually searched to check if any search terms were missing from the search strategy. The final strategy (S1) was used in the PubMed and Embase databases in August 2017.

Inclusion criteria

We included all papers with the main aim of developing, validating or updating a model predicting the risk of endometrial cancer in the general population or presence in symptomatic women. Any multivariable (at least two predictors) prediction model was eligible for inclusion, including prediction scores or prediction tools. Only papers written in the English language were included. There was no restriction on publication date.

Screening process and data extraction

Two authors performed the screening process and data extraction. One author (MA) reviewed the titles and abstracts of all papers that were identified during the search, after which a random sample of 10% was checked by another author (KV). Both authors independently screened the full text of the remaining papers for eligibility. Disagreements were solved by discussion between the authors or consulting a senior author (YV). The data extraction sheet was based on the CHARMS checklist. The data extraction sheet was pilot tested on two articles to ensure consistency between both authors. Subsequently, both authors performed the data extraction on all included papers. Specific attention was paid to four main topics (study design and methods, outcome and predictors, model development, model performance and model validation) of the CHARMS checklist, as these topics mainly influence the validity of the models. Study design and methods: We identified the study design (e.g. casecontrol, cohort, case-cohort), source of data (e.g. hospital based or national registries) and size of the study population. In addition, the inclusion criteria for each study were assessed. Outcome and predictors: We assessed the measurement and definition of both the outcome and predictors, and the handling of predictors (e.g. predictors were kept continuous or were dichotomized). Model development: We assessed the following topics: handling of predictors, number of events per variable (EPV), number and handling of missing data (e.g. single imputation, multiple imputation), methods for selection of predictors in the multivariable model (e.g. univariate analyses or subject matter knowledge) and during multivariable modelling (backward or forward selection), modelling method (e.g. logistic regression, cox proportional hazards), shrinkage (e.g. penalized shrinkage or lasso) and model presentation (e.g. regression formula, score chart, nomogram or risk score). Model performance and validation: Aspects concerning model performance and validation that were assessed were discrimination, calibration, internal validation (e.g. split-sample approach, cross-validation, bootstrapping) and external validation (e.g. geographical or temporal validation). Furthermore, the sensitivity, specificity, positive and negative likelihood ratios, and positive and negative predictive values of the diagnostic models were included in this topic, if reported.

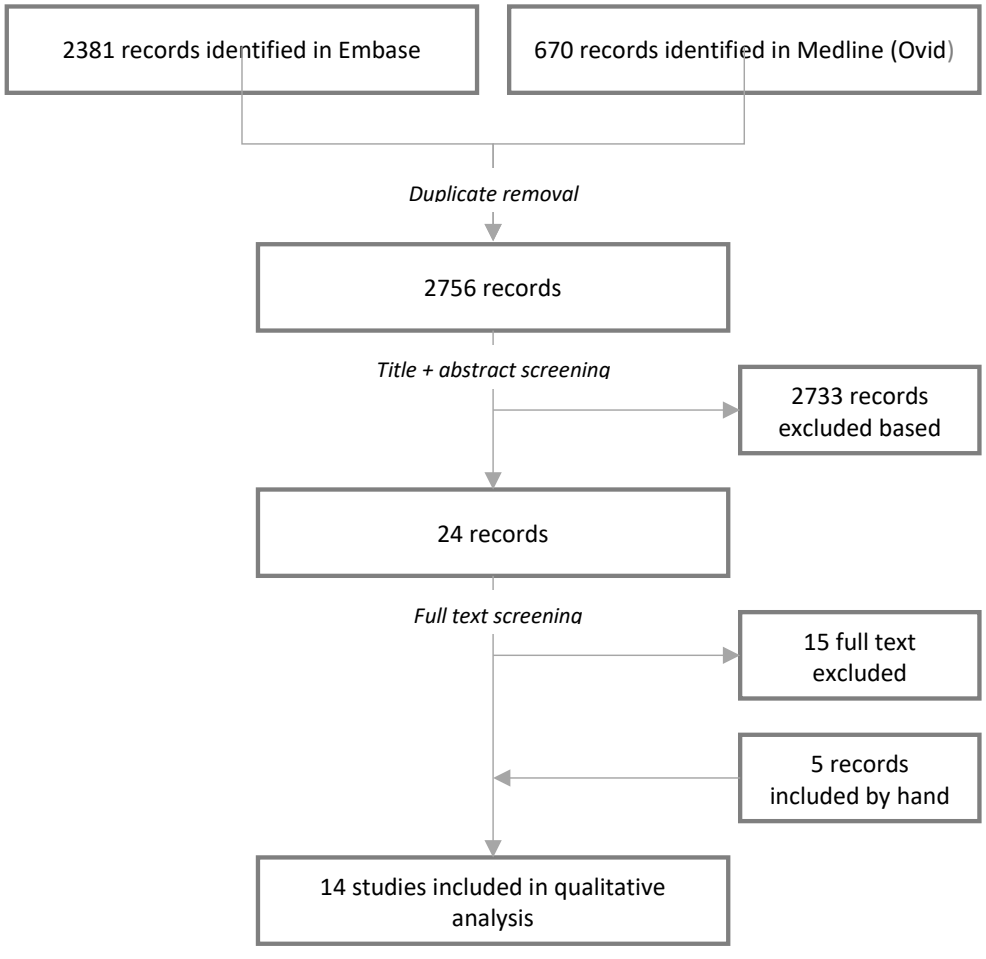

Figure 1 - Flow diagram of study selection process

Table 1: Characteristics of 12 included studies

First author	Year	Study design	Recruitment period	Total number of participants (number of cases)	Inclusion criteria	Aim of study
Risk models						
Fortner	2017	Nested case-cohort	1992-2000	716 (247)	Women without hysterectomy and prevalent cancer	Extension
Hüsing	2015	Prospective cohort	1992-2000	201,811 (855)	Women without hysterectomy and prevalent cancer	Development
Pfeiffer	2013	Prospective cohort	PLCO: 1993-2001 NIH-AARP: 1995-1996 NHS: 1990-2004	41,694 (462) 104,985 (1,097) 37,241 (532)	Non-Hispanic white women aged 50+	Development + validation
Diagnostic models						
Madkour	2017	Prospective cohort	Unknown	60 (10)	Women with postmenopausal bleeding	Development
Plotti	2017	Prospective cohort	2013-2016	298 (102)	Women aged 30 to 80 years with ultrasound endometrial abnormalities and scheduled hysteroscopy	Validation
Wong	2016	Retrospective cohort	2002-2013	4383 (168)	Women with vaginal bleeding	Development
Sladkevicius	2016	Prospective cohort	2009-2014	350 (80)	Postmenopausal women presenting with vaginal bleeding, ET ≥ 4.5 mm, and no fluid in the uterine cavity	Validation
Giannella	2014	Prospective observational	2008-2013	624 (72)	Postmenopausal women presenting with vaginal bleeding and ET ≥ 4 mm undergoing diagnostic hysteroscopy	Development
Angioli	2013	Prospective cohort	2010-2012	675 (88)	Women aged 40 to 65 years with ultrasound endometrial abnormalities and scheduled hysteroscopy	Development
Opolskiene	2011	Prospective cohort	2002-2009	261 (63)	Postmenopausal women presenting with vaginal bleeding, ET ≥ 4.5 mm and no fluid in the uterine cavity	Development
Burbos	2011	Prospective cohort	2006-2009	3347 (201)	Postmenopausal women presenting with vaginal bleeding	Development
Musonda	2011	Prospective cohort	2006-2009	3795 (221)	Postmenopausal women presenting with vaginal bleeding	Comparison
Burbos	2010	Prospective cohort	2006-2009	3047 (149)	Postmenopausal women presenting with vaginal bleeding	Development
Weber	1999	Case-control	1993-1995	194 (57)	Perimenopausal or postmenopausal women presenting with vaginal bleeding who had endometrial sampling	Development

RESULTS

We identified 2756 papers during the initial search. These records were screened on title and abstract after which 23 records were included for full text screening. The low sensitivity of the search (less than 1% of the initial search result was included for full text screening) is in line with other searches, as a consequence of the lack of adequate search terms for prediction models. After full text screening, 9 papers were eligible for inclusion. In addition, 5 extra papers were identified by hand search, leading to the inclusion of 14 papers in this review (Fig. 1). Two papers developed prediction models for the general population (risk models), eight papers developed prediction models for symptomatic women (diagnostic models), one paper internally evaluated a model, two papers described the external validation of previous developed models and one paper described the extension of an existing prediction model.

Prediction models for endometrial cancer in the general population (risk models)

Study designs and population

The two studies that developed risk models used data from population based cohorts; one study used the European EPIC cohort (Husing et al., 2016) and one study used a cohort from the United States (Pfeiffer et al., 2013) (US) (Table 1). The data was collected using a prospective cohort design. The sizes of the study populations were large: 146,679 (1559 with cancer) (Pfeiffer et al., 2013) and 201,811 (855 with cancer) (Husing et al., 2016). Inclusion criteria were similar for the two studies.

Outcome and predictors

Cases were identified through record linkage with regional cancer registries, linkage to health insurance records, active follow-up of study subjects and systematic requests of patient records from pathology registries in the European study (Husing et al., 2016). Identification of cases in the other study was done via linkage with state cancer registries, annual study updates and reviews of medical records (Pfeiffer et al., 2013). Both studies included predictors that were previously associated with endometrial cancer. The European study handled most of the continuous predictors as a linear term and centered them at the median (Husing et al., 2016), while the other study categorized all continuous predictors (Pfeiffer et al., 2013). Both studies investigated interaction effects between predictors and had an EPV above ten.

Model development

Both studies encountered incomplete data. The US study created an indicator variable for the predictor (benign breast disease) with 20% missing data and excluded all women with missing data for other predictors (Pfeiffer et al., 2013). The other study used a single,

simple imputation for the missing data (Husing et al., 2016). Both prediction models were developed with Cox proportional hazard regression for the relative risks with an additional cause specific competing risk analysis to enable computation of the age specific absolute risk of developing endometrial cancer. The two studies used a stepwise backward selection procedure to identify the strongest predictors, with alpha 0.01 (Pfeiffer et al., 2013) and 0.1 (Husing et al., 2016). The developed models both included the predictors age at menopause, BMI, parity, (duration of) oral contraceptives and menopausal hormone therapy (MHT) use, and smoking. The European model also included age at menarche, age at first full term pregnancy, and an interaction term between age at menarche and BMI. The US model included an interaction term of MHT use and BMI < 25 kg/m2 (Table 2). The European study used a bootstrap sample procedure followed by linear shrinkage to improve internal validity (Husing et al., 2016).

Model performance and validation

The US study used independent data from the Nurses' Health Study (NHS) to assess model performance. The validation dataset consisted of 37,241 participants and 532 cases. Performance of the risk model in the external dataset was assessed with the c-statistic (0.68 [0.66–0.70]) and expected versus observed ratio (E/O ratio) (1.20 [1.11–1.30]) (Table 3). The European study used five-fold cross validation; no specification was given about the split of the data. The discriminative ability was assessed with the c-statistic (0.77 [0.68–0.85]) and the calibration with the E/O ratio (0.99) and the Hosmer-Lemeshow test (p = 0.08). The integrated discrimination index (IDI) was also estimated (0.18% [0.04- 0.3]) to examine the difference between the developed model and a model that only included age and country (Table 3). The European model was extended with serum-based biomarkers in a separate study (Fortner et al., 2017). The same population (EPIC cohort) was used for updating, with a nested case-cohort design (716 participants, 247 cases). The biomarkers include adiponectin, total cholesterol, HDL cholesterol, C-peptide, C-reactive protein, androstenedione, DHEAS, oestrone, glucose, IGFBP1, IGFBP2, IL1Ra, IL6, SHBG, testosterone, TNF receptor 1, TNF receptor 2, TNFα and triglycerides. The biomarkers were log2-transformed and adjusted for age, center and menopausal status. Missing data was imputed with mean values. A backward selection procedure with alpha 0.157 was used to select serum-based biomarkers. The improvement of the model was assessed with the c-statistic and showed 0.02 points improvement (from 0.627 to 0.647, corrected for optimism) for the model with all biomarkers, and 0.017 points improvement (from 0.627 to 0.644, corrected for optimism) for the model with selected biomarkers (Table 3).

Table 2: Overview of predictors included in risk models and diagnostic models for endometrial cancer

	Age	Age at menopause	Age at menarche	BMI	(Duration of) HRT use	(Duration of) OC use	Parity	Age at FFTP	Menopausal status	Smoking	Biomarkers[2]	Recurrent bleeding	Endometrial thickness	Hypertension	Diabetes	VAS	Vascularity	Use of warfarin	Ill-defined endometrium-myometrium interface	Irregular endometrial midline	Heterogeneous endometrium	Presence of symptoms	HE4 levels
jRisk models																							
Fortner et al. (2016)	X	X	X	X	X	X	X	X	X	X	X												
Hüsing et al. (2015)	X	X	X	X	X	X	X	X	X	X													
Pfeiffer et al. (2013)	X	X		X	X	X	X		X	X													
Diagnostic models																							
Madkour (2017)													X				X		X	X	X		
Wong et al. (2016)	X	X		X			X					X											
	X						X					X	X										
Giannella et al.(2014)	X											X	X	X									
Angioli et al. (2013)	X												X									X	X
Opolskiene et al.(2011)	X				X													X					
	X												X					X					
	X				X								X			X							
	X				X								X				X						
Burbos et al. (2011)	X			X								X			X								
Burbos et al. (2010)	X			X								X	X		X								
Weber et al. (1999)	X			X[1]	X	X	X		X					X	X								
				X[1]			X								X								
							X		X						X								

Table 3: Model performance and validation measures

	E/O ratio	Hosmer-Lemeshow test	c-statistic	Internal validation	External validation
Risk models					
Fortner et al. (2017)			0.64 [0.60-0.69]	X	
Hüsing et al. (2015)	X	X	0.77 [0.68-0.85]	X	
Pfeiffer et al. (2013)	X		0.68 [0.66-0.70]		X
Diagnostic models					
Madkour (2017)			0.95		
Wong et al. (2016)			0.71 [0.66-0.75] 0.93 [0.90-0.95]	X X	
Giannella et al.(2014)			0.88 [0.84-0.91]	X	
Angioli et al. (2013)			0.957 [0.91-0.98]	X	X[2]
Opolskiene et al.(2011)			0.74 [0.67-0.81] 0.82 [0.76-0.87] 0.89 [0.84-0.94] 0.91 [0.87-0.95]		 X[1] X[1]
Burbos et al. (2011)			0.73 [0.70-0.77]		
Burbos et al. (2010)			0.77	X	
Weber et al. (1999)			0.75 0.74 0.66	X X X	

[1] = external validation was performed by Sladkevicius et al.

[2] = external validation was performed by Plotti et al.

Prediction models for endometrial cancer for symptomatic women (diagnostic models)

Study designs and population

Seven out of eight studies developed diagnostic models in cohorts from Europe (n = 5), the Middle East (n = 1) or Hong Kong (n = 1). One study used a case-control design for model development with participants from the US (Weber et al., 1999) (Table 1). Population size varied considerably between studies, from 60 (10 cases) to 4383 (168 cases). Seven studies included women presenting with postmenopausal bleeding. Postmenopausal bleeding was in four studies defined as vaginal bleeding after at least one year of spontaneous amenorrhoea (Burbos et al., 2010, 2011; Opolskiene et al., 2011; Sladkevicius and Valentin, 2016). Two studies extended this definition with a minimum age of 40 years (Giannella et al., 2014; Madkour, 2017) and one study did not specify any definition (Wong et al., 2016). Two

studies only included women with postmenopausal bleeding and endometrial thickness larger than 4 mm (Giannella et al., 2014) or 4.5 mm (Opolskiene et al., 2011). One study included women aged 40–65 years with ultrasound endometrial abnormalities (endometrial thickness, polyps and submucous myoma) and scheduled hysteroscopy (Angioli et al., 2013).

Outcome and predictors

The diagnosis of endometrial cancer in all studies was based on histopathology of the tissue that was obtained during endometrial sampling. Two studies restricted endometrial sampling to women with endometrial thickness over 5 mm, based on transvaginal ultrasonic scanning (Burbos et al., 2010, 2011). All studies included age as predictor in the multivariable modelling, complemented with epidemiological predictors, such as BMI and parity, or clinical predictors such as endometrial thickness and the presence of hypertension. One study also included HE4 (Human Epididymis Protein 4) levels, a tumour marker that can be obtained via a blood test. Six out of eight studies had an EPV of more than ten, the other studies had an EPV of 8 (Weber et al., 1999) and below 1 (Madkour, 2017). Continuous predictors were kept continuous in seven studies (Wong et al., 2016; Burbos et al., 2010; Giannella et al., 2014; Weber et al., 1999; Burbos et al., 2011; Opolskiene et al., 2011; Angioli et al., 2013), one study dichotomized the continuous predictors (Madkour, 2017). 3.2.3.

Model development

Five studies (Wong et al., 2016; Burbos et al., 2010; Weber et al., 1999; Burbos et al., 2011; Opolskiene et al., 2011) mentioned the presence of missing data, but the handling of missing data was described in only one study (Wong et al., 2016). Wong et al. imputed the mean for some variables and assigned the category 'multiparous' to women with missing data for the variable parity, as preliminary analysis showed no significance for this missing category. All studies used logistic regression analyses for model development. Six studies performed a preselection of predictors based on univariable analyses, with a p-value of 0.20 (Wong et al., 2016; Weber et al., 1999) or p-value of 0.05 (Giannella et al., 2014; Opolskiene et al., 2011; Madkour, 2017; Angioli et al., 2013). The methods and criteria for selection of predictors during multivariable modelling varied among the studies. Forward selection, backward selection and full model approaches were used. Selection was based on p-values of 0.05 in five studies (Wong et al., 2016; Giannella et al., 2014; Weber et al., 1999; Opolskiene et al., 2011; Angioli et al., 2013). One study did not clearly report the methods that were used during multivariable modelling (Madkour, 2017). The predictors that were included in the developed models varied, ranging from models with mostly epidemiological predictors, to models that only included predictors related to abnormalities of the endometrium (Table 2).

Model performance and validation

Performance of the diagnostic models was assessed with measures specific for diagnostic tests such as sensitivity and specificity (n = 7), positive and negative likelihood ratio

(n = 6), positive and negative predictive value (n = 5) and the Youden index (n = 2). All studies used the area under the ROC curve (AUC), which is the same as the c-statistic in logistic regression analysis, to describe the discriminative ability of the prediction model. The AUC varied from 0.73 to 0.957. The goodness-of-fit was assessed with the Hosmer-Lemeshow test in two studiesx (Giannella et al., 2014; Weber et al., 1999). Three studies (Wong et al., 2016; Giannella et al., 2014; Angioli et al., 2013) used internal validation with a split-sample method, with a split into two equal parts (Giannella et al., 2014) or in three parts (Wong et al., 2016; Angioli et al., 2013) (two third for model development, one third for validation). One study (Sladkevicius and Valentin, 2016) assessed the added value of endometrial thickness to a model that included only clinical predictors. Two external validation studies of diagnostic models were found. Both studies were temporal validations, as data for model development and external validation was collected in the same hospital but data used in the validation was more recent. The authors of the external validation studies were involved in the original model development. Both studies used a prospective cohort to validate the developed model with sample sizes of 80 women with cancer (Sladkevicius and Valentin, 2016) and 102 women with cancer (Plotti et al., 2017). One study described the performance of the models in the external data with the AUC, which was comparable to the performance in the original data(AUC of 0.89 and 0.91 (0.91 and 0.89 in the original data respectively)) (Sladkevicius and Valentin, 2016). The calibration was shown with calibration plots, without further specification of the intercept and slope. One study described the model performance with the predicted and observed number of malignant cases (93 predicted versus 102 observed) and benign cases (187 predicted versus 196 observed) and used a predefined cut-off point to determine the sensitivity, specificity, and the positive and negative predictive value. The model showed improved performance in terms of sensitivity (94% versus 89%) and positive predictive value (0.91 versus 0.73).

DISCUSSION

This review shows the reported prediction models for risk of endometrial cancer in the general population and presence of endometrial cancer in symptomatic women. Most models were developed and validated in European and Northern American populations. One of the two risk models was externally validated in a large independent cohort (Pfeiffer et al., 2013). Only three of the fourteen diagnostic models were externally validated in a temporal validation study (Sladkevicius and Valentin, 2016; Plotti et al., 2017). No head to head comparison of developed models was found.

Age was included in almost all models; BMI and parity were also frequently included. The risk models further included variables related to the reproductive history of women, hormone use, and smoking. All these predictors are known for the relation with developing

endometrial cancer (Smith et al., 2003; Beral et al., 2005; Lesko et al., 1985; Hinkula et al., 2002; Kitson et al., 2017). The relation between age and the development of several types of cancer has been most extensively described (DePinho, 2000; Balducci and Ershler, 2005; Anisimov, 2003). In the diagnostic models clinical predictors were more important, such as endometrial thickness, recurrent bleeding and insulin resistance.

Shortcomings in methodology were found in almost all studies, which is consistent with the findings from other systematic reviews on prediction models in different research areas (Damen et al., 2016; Mushkudiani et al., 2008; Collins et al., 2013; Bouwmeester et al., 2012). Information on handling of this missing data was limited. Nine (Pfeiffer et al., 2013; Wong et al., 2016; Husing et al., 2016; Burbos et al., 2010; Fortner et al., 2017; Weber et al., 1999; Burbos et al., 2011; Opolskiene et al., 2011; Sladkevicius and Valentin, 2016) out of fourteen studies mentioned the presence of missing data, of which four studies (Pfeiffer et al., 2013; Wong et al., 2016; Husing et al., 2016; Fortner et al., 2017) described the handling of the missing data. Handling varied from creating an indicator variable for the missing data to single, simple imputation. A previous simulation study has shown that the indicator method will lead to biased results, even when the missing data is missing completely at random (Donders et al., 2006). None of the models used multiple imputation, while this is considered the preferred method (Donders et al., 2006; Marshall et al., 2010).

The way of identifying the strongest predictors varied between the models. Both forward and backward procedures were used with different stopping rules. The choice for a stopping rule has an influence on the number of predictors that will be selected during the modelling process, which may result in overfitting depending on the sample size (Steyerberg, 2009). Six out of eight studies with diagnostic models (Wong et al., 2016; Giannella et al., 2014; Weber et al., 1999; Opolskiene et al., 2011; Madkour, 2017; Angioli et al., 2013) used a rather stringent p-value of 0.05 during multivariable modelling, while samples sizes were relatively small. Two of these diagnostic models had an EPV below ten, which might contribute to overfitting of the prediction model, resulting in too extreme predictions for new patients. Especially the performance (c-statistic of 0.95) of the model with EPV below 1 might be too optimistic, and will probably deteriorate during external validation.

Reproducibility of the results can be studied with internal validation (Steyerberg and Vergouwe, 2014). Bootstrapping is considered the most efficient method to study internal validity in small datasets (Moons et al., 2012). Three studies (Husing et al., 2016; Fortner et al., 2017; Weber et al., 1999) used bootstrapping to internally validate the developed models and three studies (Wong et al., 2016; Giannella et al., 2014; Angioli et al., 2013) used a split-sample. Remarkable is the use of bootstrapping in the EPIC study, as their sample size was large (855 women with cancer) and bootstrapping might therefore be considered

unnecessary as overfitting and optimism is limited in such large sample sizes (Steyerberg, 2009).

It is important to assess the validity of a model in independent data from a different setting (geographical validation) or from a more recent time period (temporal validation) (Steyerberg, 2009; Altman and Royston, 2000; Justice et al., 1999). Three studies performed an external validation, in which the models showed relatively good performance in the independent data (Pfeiffer et al., 2013; Sladkevicius and Valentin, 2016; Plotti et al., 2017). One model (Pfeiffer et al., 2013) overestimated the number of cases, indicating suboptimal calibration in the large. This was related to the difference in average risk in the population that was used for model development. Miscalibration in the large is often found and can easily be adjusted for (te Velde et al., 2014; Vergouwe et al., 2010).

The two risk models included in this review have moderate discriminative ability, which did not improve substantially by updating the model with serum-based biomarkers. This implies that new information may be needed to improve the model performance. Improvement is necessary because implementing a model with poor discriminative ability has little value in practice. New information can for example be found in the field of epigenetics. Current developments in other research areas have already shown that the epigenome contains objective information on environmental exposure and can be useful for making risk predictions. Alterations in DNA methylation were shown to be the consequence of adiposity. The change in DNA methylation predicted the risk of developing type 2 diabetes (Wahl et al., 2017). In addition, the hypomethylation of the aryl hydrocarbon receptor repressor (AHRR) gene holds information on former smoking status, which might contribute to risk predictions of lung cancer (Bojesen et al., 2017). Further, the methylation of peripheral blood cell DNA can serve as a predictor for the risk of developing breast cancer (Widschwendter et al., 2008). Adding information from the epigenome to existing risk prediction models likely improves their performance, as information from the epigenome is less error-prone and may therefore hold more information than data from questionnaires (Ladd-Acosta and Fallin, 2016; Pashayan et al., 2016). Adding genetic information for instance on mutations in mismatch repair genes, as seen in Lynch syndrome, may also improve model performance. Women with Lynch syndrome have approximately a 20–60% cumulative lifetime risk of developing endometrial cancer and Lynch syndrome is responsible for 2-5% of all endometrial cancer cases (Meyer et al., 2009).

The total number of studies included in this review is relatively small, especially the number of models that predict the risk of developing endometrial cancer in the general population. Despite extensive searches no additional studies were found. The small number of studies may have an influence on the overview of selected predictors, as more studies might give a more complete and representative reflection of important predictors.

In conclusion, only a few models have been developed to predict endometrial cancer in asymptomatic or symptomatic women. The usefulness of most of the models is unclear considering methodological shortcomings and lack of external validation and head to head comparisons of models. Developed risk models should be externally validated and extended with new predictors, such as genetic and epigenetic risk predictors, to improve model performance. Future research on diagnostic models should focus on external validation and creating models with larger sample sizes, which could be realized with individual patients data meta-analysis.

REFERENCES

1. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur J Cancer. 2013;49:1374-403.
2. Kyrgiou M, Kalliala I, Markozannes G, Gunter MJ, Paraskevaidis E, Gabra H, et al. Adiposity and cancer at major anatomical sites: umbrella review of the literature. Bmj. 2017;356:j477.
3. Dixon JB. The effect of obesity on health outcomes. Mol Cell Endocrinol. 2010;316:104-8.
4. Collaboration NCDRF. Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants. Lancet. 2016;387:1377-96.
5. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2014;384:766-81.
6. MacMahon B. Risk factors for endometrial cancer. Gynecologic Oncology. 1974;2:122-9.
7. Hecht JL, Mutter GL. Molecular and Pathologic Aspects of Endometrial Carcinogenesis. Journal of Clinical Oncology. 2006;24:4783-91.
8. Pfeiffer RM, Park Y, Kreimer AR, Lacey JV, Jr., Pee D, Greenlee RT, et al. Risk prediction for breast, endometrial, and ovarian cancer in white women aged 50 y or older: derivation and validation from population-based cohort studies. PLoS Med. 2013;10:e1001492.
9. Wong AS, Cheung CW, Fung LW, Lao TT, Mol BW, Sahota DS. Development and validation of prediction models for endometrial cancer in postmenopausal bleeding. Eur J Obstet Gynecol Reprod Biol. 2016;203:220-4.
10. Husing A, Dossus L, Ferrari P, Tjonneland A, Hansen L, Fagherazzi G, et al. An epidemiological model for prediction of endometrial cancer risk in Europe. Eur J Epidemiol. 2016;31:51-60.
11. Burbos N, Musonda P, Giarenis I, Shiner AM, Giamougiannis P, Morris EP, et al. Predicting the risk of endometrial cancer in postmenopausal women presenting with vaginal bleeding: the Norwich DEFAB risk assessment tool. Br J Cancer. 2010;102:1201-6.
12. Giannella L, Mfuta K, Setti T, Cerami LB, Bergamini E, Boselli F. A risk-scoring model for the prediction of endometrial cancer among symptomatic postmenopausal women with endometrial thickness > 4 mm. Biomed Res Int. 2014;2014:130569.
13. Gull B, Karlsson B, Milsom I, Granberg S. Can ultrasound replace dilation and curettage? A longitudinal evaluation of postmenopausal bleeding and transvaginal sonographic measurement of the endometrium as predictors of endometrial cancer. American Journal of Obstetrics and Gynecology. 2003;188:401-8.
14. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. Bmj. 2016;353:i2416.
15. Ingui BJ, Rogers MAM. Searching for Clinical Prediction Rules in Medline. Journal of the American Medical Informatics Association : JAMIA. 2001;8:391-7.

16. Smit HA, Pinart M, Anto JM, Keil T, Bousquet J, Carlsen KH, et al. Childhood asthma prediction models: a systematic review. Lancet Respir Med. 2015;3:973-84.
17. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. Breast Cancer Res Treat. 2012;132:365-77.
18. Mushkudiani NA, Hukkelhoven CW, Hernandez AV, Murray GD, Choi SC, Maas AI, et al. A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes. J Clin Epidemiol. 2008;61:331-43.
19. Fortner RT, Husing A, Kuhn T, Konar M, Overvad K, Tjonneland A, et al. Endometrial cancer risk prediction including serum-based biomarkers: results from the EPIC cohort. Int J Cancer. 2017;140:1317-23.
20. Weber AM, Belinson JL, Piedmonte MR. Risk Factors for Endometrial Hyperplasia and Cancer Among Women With Abnormal Bleeding. Obstetrics & Gynecology. 1999;93:594-8.
21. Burbos N, Musonda P, Duncan TJ, Crocker SG, Morris EP, Nieto JJ. Estimating the risk of endometrial cancer in symptomatic postmenopausal women: a novel clinical prediction model based on patients' characteristics. Int J Gynecol Cancer. 2011;21:500-6.
22. Opolskiene G, Sladkevicius P, Valentin L. Prediction of endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness >/= 4.5 mm. Ultrasound Obstet Gynecol. 2011;37:232-40.
23. Sladkevicius P, Valentin L. Prospective validation of two mathematical models to calculate the risk of endometrial malignancy in patients with postmenopausal bleeding and sonographic endometrial thickness >/=4.5 mm. Eur J Cancer. 2016;59:179-88.
24. Madkour NM. An ultrasound risk-scoring model for prediction of endometrial cancer in post-menopausal women (using IETA terminology). Middle East Fertility Society Journal. 2017;22:201-5.
25. Angioli R, Capriglione S, Aloisi A, Luvero D, Cafa EV, Dugo N, et al. REM (risk of endometrial malignancy): a proposal for a new scoring system to evaluate risk of endometrial malignancy. Clin Cancer Res. 2013;19:5733-9.
26. Plotti F, Capriglione S, Terranova C, Montera R, Scaletta G, Lopez S, et al. Validation of REM score to predict endometrial cancer in patients with ultrasound endometrial abnormalities: results of a new independent dataset. Med Oncol. 2017;34:82.
27. Smith JS, Green J, Berrington de Gonzalez A, Appleby P, Peto J, Plummer M, et al. Cervical cancer and use of hormonal contraceptives: a systematic review. Lancet. 2003;361:1159-67.
28. Beral V, Bull D, Reeves G, Million Women Study C. Endometrial cancer and hormone-replacement therapy in the Million Women Study. Lancet. 2005;365:1543-51.
29. Lesko SM, Rosenberg L, Kaufman DW, Helmrich SP, Miller DR, Strom B, et al. Cigarette smoking and the risk of endometrial cancer. N Engl J Med. 1985;313:593-6.
30. Hinkula M, Pukkala E, Kyyronen P, Kauppila A. Grand multiparity and incidence of endometrial cancer: a population-based study in Finland. Int J Cancer. 2002;98:912-5.
31. Kitson SJ, Evans DG, Crosbie EJ. Identifying High-Risk Women for Endometrial Cancer Prevention Strategies: Proposal of an Endometrial Cancer Risk Prediction Model. Cancer Prev Res (Phila). 2017;10:1-13.

32. DePinho RA. The age of cancer. Nature. 2000;408:248-54.
33. Balducci L, Ershler WB. Cancer and ageing: a nexus at several levels. Nat Rev Cancer. 2005;5:655-62.
34. Anisimov VN. The relationship between aging and carcinogenesis: a critical appraisal. Crit Rev Oncol Hematol. 2003;45:277-304.
35. Collins GS, Omar O, Shanyinde M, Yu LM. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. J Clin Epidemiol. 2013;66:268-77.
36. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med. 2012;9:1-12.
37. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol. 2006;59:1087-91.
38. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. BMC Med Res Methodol. 2010;10:7.
39. Steyerberg EW. Clinical Prediction Models: Springer-Verlag New York; 2009.
40. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35:1925-31.
41. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio) marker. Heart. 2012;98:683-90.
42. Altman DG, Royston P. What do we mean by validating a prognostic model? Statistics in Medicine. 2000;19:453-73.
43. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med. 1999;130:515-24.
44. te Velde ER, Nieboer D, Lintsen AM, Braat DD, Eijkemans MJ, Habbema JD, et al. Comparison of two models predicting IVF success; the effect of time trends on model performance. Hum Reprod. 2014;29:57-64.
45.] Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. Am J Epidemiol. 2010;172:971-80.
46. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2017;541:81-6.
47. Bojesen SE, Timpson N, Relton C, Davey Smith G, Nordestgaard BG. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. Thorax. 2017;72:646-53.
48. Widschwendter M, Apostolidou S, Raum E, Rothenbacher D, Fiegl H, Menon U, et al. Epigenotyping in peripheral blood cell DNA and breast cancer risk: a proof of principle study. PLoS One. 2008;3:e2656.
49. Ladd-Acosta C, Fallin MD. The role of epigenetics in genetic and environmental epidemiology. Epigenomics. 2016;8:271-83.
50. Pashayan N, Reisel D, Widschwendter M. Integration of genetic and epigenetic markers for risk stratification: opportunities and challenges. Per Med. 2016;13:93-5.

51. Meyer LA, Broaddus RR, Lu KH. Endometrial cancer and Lynch syndrome: clinical and pathologic considerations. Cancer Control. 2009;16:14-22.

Chapter 5

Development and Validation of a Model to Predict Regression of Large Size Hepatocellular Adenoma

Anne Julia Klompenhouwer, Maaike Alblas, Belle Vivica van Rosmalen, Martinus Petrus Daniel Haring, Esmee Venema, Michail Doukas, Maarten Guillaume Josephus Thomeer, Robert Bart Takkenberg, Joanne Verheij, Vincent Erwin de Meijer, Thomas Matthijs van Gulik, Hester Floor Lingsma, Robert Auke de Man, and Jan Nicolaas Maria Ijzermans,

American Journal of Gastroenterology. 2019

ABSTRACT

Introduction: Surgery is advocated in hepatocellular adenomas (HCA) >5 cm that do not regress to <5 cm after 6–12months. The aim of this study was to develop a model for these patients, estimating the probability of HCA regression to <5 cm at 1 and 2 years follow-up.

Methods: Data were derived from a multicenter retrospective cohort of female patients diagnosed with HCA >5 cm at first follow-up. Potential predictors included age, body mass index, and HCA diameter at diagnosis (T0), HCA-subtype (hepatocyte nuclear factor 1a inactivated HCA, inflammatory-HCA, unclassified HCA) and "T0-T1 regression-over-time" (percentage of regression between T0 and first follow-up (T1) divided by weeks between T0 and T1). Cox proportional hazards regression was used to develop a multivariable model with time to regression of HCA < 5 cm as outcome. Probabilities at 1 and 2 years follow-up were calculated.

Results: In total, 180 female patients were included. Median HCA diameter at T0 was 82.0 mm and at T1 65.0 mm. Eighty-one patients (45%) reached the clinical endpoint of regression to <5 cm after a median of 34 months. No complications occurred during follow-up. In multivariable analysis, the strongest predictors for regression to <5 cm were HCA diameter at T0 (log transformed, hazard ratio (HR) 0.05), T0-T1 regression-over-time (HR 2.15) and HCA subtype inflammatory-HCA (HR 2.93) and unclassified HCA (HR 2.40), compared to hepatocyte nuclear factor 1a inactivated HCA (reference). The model yielded an internally validated c-index of 0.79.

Conclusions: In patients diagnosed with HCA >5 cm that still exceed 5 cm at first follow-up, regression to <5 cm can be predicted at 1 and 2 years follow-up using this model. Although external validation in an independent population is required, this model may aid in decision-making and potentially avoid unnecessary surgery.

INTRODUCTION

Hepatocellular adenoma (HCA) is a rare benign liver tumor that is usually discovered incidentally in women using estrogen containing oral contraceptives (OC). It has been associated with obesity, metabolic disorders, and the intake of androgens. With cessation of OC and weight reduction regression of HCA may occur (1–3). HCA can be subdivided based on genetic and phenotypic characteristics, among which Hepatocyte Nuclear Factor 1a inactivated (H-HCA), inflammatory-HCA (I-HCA), b-catenin-activated (b-HCA), b-catenin-activated inflammatory (b-IHCA), and most recently, sonic hedgehog (sh-HCA) adenomas (4,5) (Table 1). HCA with no specific mutations are termed unclassified adenomas (U-HCA). These subtypes may be distinguished radiologically or based on immunohistochemical staining or molecular characterization. Contrast-enhanced magnetic resonance imaging (MRI) has the highest sensitivity and specificity for diagnosis of HCA and may also be used for subtype determination (6,7). Liver biopsy can be performed in the case of inconclusive imaging or when its result is expected to impact treatment decisions.

Table 1: HCA subtypes

H-HCA	Inactivating mutation of Hepatocyte Nuclear Factor 1α
I-HCA	JAK/STAT pathway activation, caused by mutations in different parts of the signaling pathway.
β-HCA	Mutation in either exon 3 or exon 7/8 of the CTNNB1 gene, causing activation of the β-catenin protein. At risk for malignant transformation.
β-IHCA	Both JAK/STAT pathway activation and a mutation in CTNNB1. At risk for malignant transformation.
sh-HCA	Activation of sonic hedgehog signaling pathway
U-HCA	Restgroup of HCA without distinctive underlying mutations or activations

The most common complication of HCA is hemorrhage, thought to occur mostly in I-HCA (8) and sh-HCA (5). A more rare complication is malignant transformation to hepatocellular carcinoma, occurring particularly in b-HCA or b-IHCA and in men with HCA (9,10). Both complications seem to occur mostly in HCA exceeding 5 cm (10,11). In the clinical practice guideline regarding the management of benign liver tumors, it is stated that a conservative approach with lifestyle adaption (cessation of OC, weight reduction) is justified in women with HCA (12). Resection of HCA is indicated in men, patients with b-HCA or b-IHCA, in case of significant growth, and when HCA diameter exceeds 5 cm 6 months after lifestyle changes. However, a recent study showed that the follow-up of potential regression could be prolonged to 12 months, and possibly even longer for large HCA (13) as these lesions will regress over time and sometimes even disappear completely (Figure 1). The present study focuses on patients diagnosed with HCA > 5 cm that still exceed 5 cm at first follow-up imaging. The aim of this study was to develop a clinical prediction model estimating the

probability of HCA regression to >5 cm at 1 and 2 years of follow-up, which can be used in timely selection of patients for surgery.

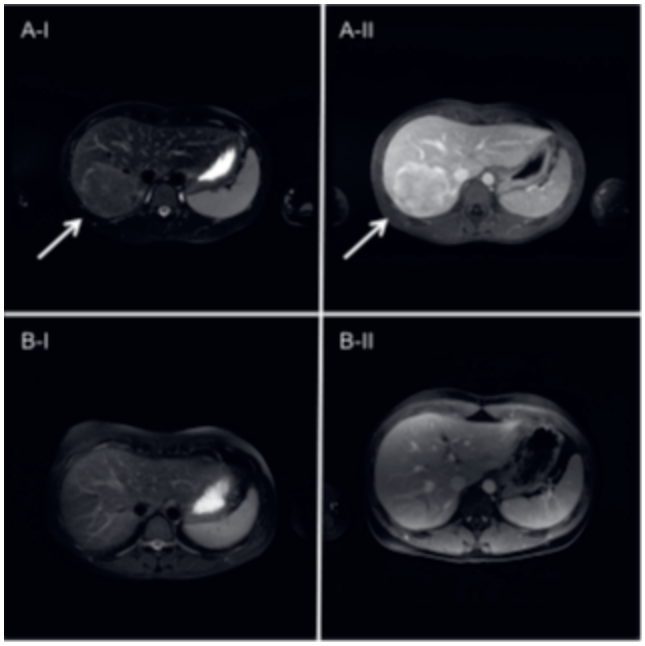

Figure 1. Example of regressing HCA. Example of a patient with a large HCA in the right hemiliver. (a) At diagnosis in 2013. A-I T2-W fatsaturated sequence. A-II T1-W fatsaturated sequence venous phase. (b) Nearly complete regression 3 years after cessation of oral contraceptives. B-I T2-W fatsaturated sequence. B-II T1-W fatsaturated sequence venous phase.

METHODS

Design and study population

Patients were derived from a retrospective cohort of patients diagnosed with HCA in 3 tertiary referral centers in the Netherlands (the Erasmus MC University Medical Center in Rotterdam, the Amsterdam University Medical Centers (location Academic Medical Center) in Amsterdam, and the University Medical Center in Groningen) between January 2000 and October 2017. HCA diagnosis was established by either contrast-enhanced MRI, histological examination (biopsy or resection specimen), or both. Patients were included if they were women and had at least 1HCA with a diameter >5 cm at the moment of diagnosis (T0) as well as at first follow-up imaging (T1). The minimum follow-up time was 6 months. Men with HCA and all patients with histologically proven b-(I)HCA were excluded because resection is recommended in these patients because of higher risk of malignant transformation. Patients who underwent an intervention before their first follow-up imaging and those who experienced hemorrhage before HCA diagnosis causing an unreliable radiological

assessment of the diameter were also excluded. The study protocol was reviewed by the accredited institutional review board; informed consent was waived.

All patients were treated according to the same treatment algorithm. At diagnosis, patients were presented at a multidisciplinary tumor board to establish a definitive diagnosis. When HCA was diagnosed, patients were urged to discontinue OC and other systemic hormonal agents and to reduce weight in case of a body mass index (BMI) > 25 kg/m2. First follow-up imaging was scheduled usually around 6–12 months after diagnosis. For each patient, all follow-up imaging was discussed at the multidisciplinary tumor board, and management (continuing follow up or intervention) was determined.

Electronic medical records were retrospectively reviewed to collect clinical data including sex, age at diagnosis, diagnostic work-up (imaging modality, biopsy), date and size of HCA at time of diagnosis (T0) and first follow-up (T1), date of last follow up imaging, management (follow-up, intervention), and HCA subtype (H-HCA, I-HCA, and U-HCA). Because sh-HCA was not described until recently, it is not included as a separate subtype in this study. HCA-subtype was determined based on typical contrast-enhanced MRI features (6,7,14), immunohistochemistry (15), or patho-molecular characterization. In all 3 centers, histologic specimens have been recently revised to determine HCA subtypes of older patients diagnosed before 2013. U-HCA were only considered to be unclassified based on patho-molecular characterization, when only imaging report was available the subtype remained undetermined (missing).

The clinical outcome was regression to <5 cm, and the date of the follow-up imaging when the HCA was seen to have regressed to <5 cm for the first time was documented. In patients with multiple lesions, the size of the largest lesion was taken because the European Association for the Study of the Liver guideline states to base management decisions on the size of the largest lesion (12). A new variable was calculated to objectify the regression-over-time between T0 and T1. First, we calculated the regression coefficient between T0 and T1: (diameter HCA T0 2 diameter HCA T1)/diameter HCA at T0. This was then standardized by dividing the regression coefficient by the number of weeks between T0 and T1. This results in a new variable called “T0T1 regression-over-time.”

Statistical analysis

Continuous variables are summarized as median and interquartile range, categorical variables as frequency (n) and percentages (%). All statistical analyses were performed using IBM SPSS software version 21.0 (Chicago, IL) or R version 3.3.3.

Differences between groups were analyzed using x2 test for categorical variables. Correlation between variables was analyzed using Pearson product-moment correlation coefficient.

Overall time-to-event analysis was performed using the Kaplan-Meier method and log-rank test with regression of the HCA to <5 cm as outcome. Patients who were treated conservatively and failed to reach this clinical endpoint were censored at the time of last follow-up imaging; patients who underwent an intervention were censored at the last imaging before intervention.

To identify predictors of HCA regression to <5 cm, a multivariable Cox proportional hazards model was developed. We only considered variables that were regarded as clinically relevant and that were easily accessible. These were age at diagnosis, BMI, HCA diameter at T0, T0T1 regression-over-time, and HCA subtype. OC use was not considered as a predictor as almost all patients used OC. We used natural logarithmic transformation to correct for nonlinearity when indicated. Multiple imputation with 5 complete datasets (R, mice package, van Buuren 2017) was applied to account for missing data for BMI (14.7%) and HCA subtype (18.9%).

The inclusion of variables into the multivariable model was assessed using a stepwise backward selection method (R, rms package, Harrell 2017) based on the Akaike information criterion. We used an internal validation procedure with bootstrap resampling with 500 replications to correct the model performance for optimism, and to compute a shrinkage factor to correct for overfitting (16). Point estimates were reported as hazard ratios (HRs) with 95% confidence intervals (95% CIs). The overall performance in terms of discriminative ability of the prediction model was measured with Harrell's concordance index (C-statistic) and corrected for optimism. A C-statistic below 0.5was considered as very poor, a C-statistic over 0.7 as good, and a C-statistic over 0.8 as strong. All tests were 2-sided and a P-value , 0.05 was considered as the level of significance.

The prediction model was developed and reported in accordance with the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis guideline (see Appendix A, Supplementary Digital Content 1, http://links.lww.com/AJG/A80) (17). The following sensitivity analyses were performed to validate the model: first with only baseline characteristics, a second in patients with pathologically proven HCA and subtypes based on patho-molecular characterization only, and a third in patients who were treated conservatively only (excluding those who underwent an intervention).

RESULTS

Clinical characteristics

A total of 180 patients met the inclusion criteria: 122 from Erasmus Medical Center, Rotterdam, 30 from Amsterdam University Medical Centers, Amsterdam, and 28 from University Medical Center, Groningen. They were all women diagnosed with HCA at a median age of 36 years and with a median BMI of 32.0 kg/m2. Almost all (95.6%) used OC.

All but one patient underwent contrast-enhanced MRI, and in 98 patients (54.4%), HCA was histologically proven. No statistically significant differences in diagnostic work-up were seen between the 3 participating centers. More than half of the study population (57.2%) had I-HCA, 15% U-HCA, and 8.9% H-HCA (Table 2).

Follow-up and primary endpoint

The median follow-up time was 24.0 months, median HCA diameter at diagnosis (T0) was 82.0 mm, and at first follow-up imaging (T1), 65.0mm(Table 2). Median time between diagnosis and first follow-up imaging was 6 months (interquartile range 5–8 months). Kaplan-Meier analysis showed 81 patients reaching the clinical endpoint of regression to,5 cm(45%) after a median of 34 months since diagnosis (95% CI 25.8–42.2 months) (Figure 2a). Subanalysis in patients who used OC showed no statistically significant difference in reaching the clinical endpoint between patients with BMI,or.30 kg/m2 (P = 0.78, Figure 2b). Most of the patients were treated conservatively (67.2%), the remaining 32.8% underwent an intervention (27.8% resection, 3.3% embolization, and 1.7% radiofrequency ablation). Of the 81 patients who reached the clinical endpoint of regression to <5 cm, 8 still underwent an intervention because of an active pregnancy wish or on patients own request. No statistically significant correlation was found between the year of diagnosis and whether an intervention was performed (r = -0.145, P = 0.053). No statistically significant differences in management were seen between the 3 participating centers (P = 0.650). HCA was confirmed in all resection specimens. No growth of HCA or complications (hemorrhage or malignant transformation) occurred during the surveillance period.

Construction of the prediction model

After stepwise backward selection based on the Akaike information criterion, the final multivariable model comprised 3 variables. These were HCA diameter at T0 (log transformed, HR 0.05), T0T1 regression-over-time (HR 2.15), and HCA subtype (HR1.00 (reference), 2.93 and 2.40 for H-HCA, I-HCA and U-HCA, resp.) (Table 3). The predicted chance (%) of HCA regression to ,5 cm within 1 and 2 years after diagnosis can be determined by:

1 year after diagnosis : P = [1 – (exp(-exp(B) x 0.063))] x 100%
2 years after diagnosis : P = [1 – (exp(-exp(B) x 0.306))] x 100%
B = [(LN(HCA diameter T0) x -2.996) + (T0T1 regression-over-time x 0.736) + ([0 if H-HCA; 1.091 if I-HCA; 0.878 if U-HCA)] + 11.749] x 0.830.

The overall predictive ability for regression to <5 cm, calculated with internally validated C-statistic (corrected for optimism), was 0.79 (95% CI 0.73–0.85).

Table 2: baseline characteristics

	Included patients with HCA N = 180
Female	180 (100%)
Median age at diagnosis (yr)	36 (29 – 45)
Median BMI (kg/m^2)	32.0 (27.4 – 35.9)
Hormone usage	
Oral contraceptives	172 (95.6%)
Never	3 (1.7%)
Steroids as medication	1 (0.6%)
Unknown	4 (2.2%)
Median follow-up time (months)	24.0 (13.0 – 49.0)
Median time between diagnosis and first follow-up imaging (months)	6.0 (5.0 – 8.0)
Median diameter of HCA at diagnosis (mm)	82.0 (65.0 – 100.0)
Median diameter of HCA at first follow-up imaging (mm)	65.0 (56.0 – 80.0)
Diagnostic work-up	
Contrast enhanced MRI	179 (99.4%)
Histologically proven	98 (54.4%)
HCA subtype	
H-HCA	16 (8.9%)
I-HCA	103 (57.2%)
U-HCA	27 (15%)
Undetermined	34 (18.9%)
Management	
Conservative	121 (67.2%)
Resection	50 (27.8%)
Embolization	6 (3.3%)
Radiofrequent Ablation	3 (1.7%)

Sensitivity analyses

Three sensitivity analyses were performed. In the first, only baseline characteristics were used, so T0T1 regression over time was discarded. This resulted in a multivariable model with HCA diameter at T0(logtransformed,HR0.1) and HCA subtype (HR1.00 (reference), 9.86 and 15.34 for H-HCA, I-HCA and U-HCA, resp.). The internally validated C-statistic (corrected for optimism) was 0.79 (95% CI 0.72–0.91). Sensitivity analysis in patients with pathologically proven HCA only (n598, of which 29 reached the clinical endpoint of regression to <5 cm) provided us with a multivariable model comprising the same 3 variables as the complete analysis with similar HRs and a C-statistic of 0.77. The third analysis was performed in patients who were treated conservatively only (n= 5121, of which, 74 reached the clinical endpoint of regression to <5 cm), also resulting in a multivariable model comprising the same 3 variables as the complete analysis and a C-statistic of 0.73.

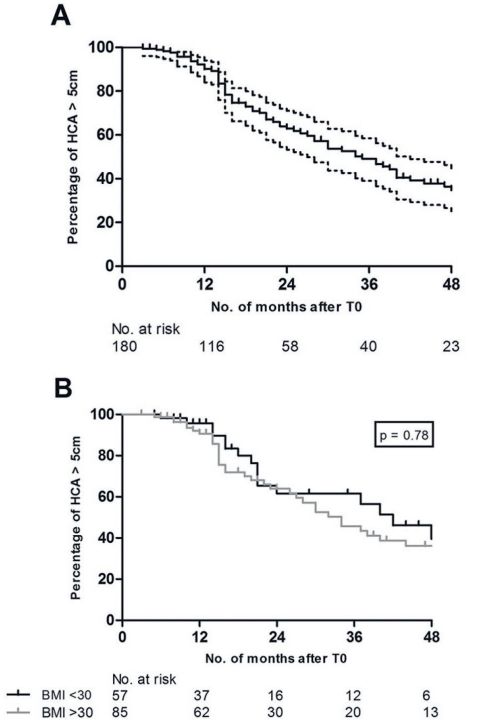

Figure 2. Kaplan-Meier analysis. (a) Kaplan-Meier curves for the event HCA regression to <5 cm in months after diagnosis (T0), median and 95% CI. (b) Subanalysis based on BMI (< or > 30 kg/m²) in patients who used oral contraceptives. BMI, body mass index; HCA, hepatocellular adenoma.

Application of the prediction model

The final model was translated into a chance assessment tool. Predictors include diameter at diagnosis, diameter at first follow-up, dates of diagnosis and first follow-up, and HCA-subtype (T0T1 regression-over-time will be calculated automatically). The chance assessment tool will provide the estimated chance of regression to <5 cm at 1 and 2 years after diagnosis (Figure 3). The chance assessment tool is available via https://hcaprediction.shinyapps.io/calculator/.

Table 3: multivariable Cox proportional hazards model

	Hazard ratio (95% confidence interval)	**p-value**
Diameter of HCA at diagnosis (logtransformed, mm)	0.05 (0.02– 0.13)	<.001
T0T1 regression over time	2.15 (1.75 – 2.70)	<.001
HCA subtype		
H-HCA	1.00 (reference)	
I-HCA	2.93 (1.19 – 7.21)	0.02
U-HCA	2.40 (0.88 – 6.55)	0.09

Predictors		Value
Diameter at diagnosis	*[mm]*	70
Date diagnosis	*[dd-mm-yyyy]*	20-09-2017
Diameter at first follow-up	*[mm]*	60
Date first follow-up	*[dd-mm-yyyy]*	01-03-2018
Subtype	*[0=H-HCA | 1=I-HCA | 2=U-HCA]*	1

Predicted chance of regression to <5cm (%)	
1 year after diagnosis	**7**
2 years after diagnosis	**29**

Figure 3. Chance calculator. An example of a patient in the chance calculator. Patient had a 70 mm inflammatory HCA at diagnosis that regressed to 60 mm at first follow-up. The predicted chance of regression to <5 cm is 7% 1 year after diagnosis and 29% 2 years after diagnosis. The chance calculator is available via https://hcaprediction.shinyapps.io/calculator/. H-HCA, HNF-1a inactivated HCA; I-HCA, inflammatory HCA; U-HCA, unclassified HCA.

DISCUSSION

In this study of 180 female patients diagnosed with HCA >5 cm in 3 tertiary referral centers in the Netherlands, we present a clinical chance assessment tool able to predict the probability of HCA regression to ,5 cm at 1 and 2 years after diagnosis. The model comprises 3 easily accessible variables: HCA diameter at diagnosis, T0T1 regression-over-time, and HCA subtype. This study is the first to develop a prediction model from a clinical perspective for patients with HCA. The model can be used for patients diagnosed with HCA >5 cm that still exceed 5 cm at first follow-up imaging and estimates the chance of regression to <5 cm at 1 and 2 years after diagnosis. It can be of aid to clinicians in decisions pertaining to surgery or continued surveillance. Using this model, resection can be reserved for patients with low probability of HCA regression, whereas unnecessary resection in patients with a high chance of HCA regression can be avoided. A considerable health benefit could be provided as HCA is associated with obesity and the complication risk following surgery is significantly increased in such patients (18,19).

To identify factors predictive of HCA regression to <5 cm, 5 variables were considered to be clinically relevant and easily accessible. Age at diagnosis and BMI turned out to be the least predictive and were therefore not included in the final model. Ideally, we would have wanted to add change in BMI as a potential predictive variable, as weight loss seems to be a factor to cause regression of HCA (2). In our series a subanalysis in patients who used OC showed no significant differences in reaching the clinical endpoint of,5 cm between patients with BMI, or. 30 kg/m2. Unfortunately, change in BMI was underreported in all centers.

Our results show that the association between a high BMI at diagnosis and regression of HCA is minor if anything, future studies should focus on prospectively assessing the association between regression and weight loss. The finding that age is not a significant predictor

surprised us, as it has been established that HCA regress after menopause (20). This may be attributed to the fact that the effect of cessation of OC causes the first regression, whereas the effect of age will not be noticeable until a later stage of follow-up.

The model shows that the size of the HCA at diagnosis and T0T1 regression-over-time, defining the regression of the HCA over time between diagnosis and first follow-up imaging, are associated with regression to <5 cm. Larger HCA have a lower chance of reaching the clinical endpoint of <5 cm, as do HCA that show little regression in the first follow-up period. Hemorrhage was only seen on diagnostic imaging and did not occur after the establishment of HCA diagnosis and cessation of OC and no malignant transformation was seen during follow-up. This suggests that a significant decrease in size as such might be just as relevant to prevent bleeding as the decrease in size to ,5 cm. Currently, there are no data supporting the concept that there is still a risk of bleeding in regressing HCA. Therefore, surgery might not be necessary even in HCA .5 cm to prevent bleeding because the risk of bleeding in a HCA showing regression in size, apparently is very small. This study supports this concept in a large clinical series.

The results show that the chance of regression to <5 cm is lower in H-HCA, compared with I-HCA and U-HCA. However, given the low risk of complications in H-HCA, a conservative approach seems justified in confirmed H-HCA, independent of the chance of regression. In addition, because this study might lead to a more conservative approach regarding HCA in general, subtype determination and biopsy within the diagnostic workup becomes increasingly important. In this study, we deliberately excluded men with HCA and patients with histologically proven b-(I)HCA because resection is recommended in these cases, given the higher risk of malignant transformation (9,10). In addition, early resection might be performed in sh-HCA, given the apparent higher risk of bleeding (5). A conservative approach may only be justified when HCA-subtype is established in I-HCA and H-HCA, preferably with biopsy.

We performed 3 different sensitivity analyses in this study. In the first, the model was developed with baseline characteristics only to see whether the model may be used at diagnosis as well. This resulted in a model with HCA diameter at diagnosis and HCA-subtype only and a comparable C-statistic as compared to the complete analysis. We believe, however, that the model is of more interest to use after the first follow-up imaging because a conservative approach with lifestyle changes and follow-up is advised in all patients with HCA, irrespective of the diameter. The second and third sensitivity analyses were performed in patients with pathologically proven HCA and those who were treated conservatively only. Both show a model comprising the same variables as the model from the complete analysis with comparable HR's. We aimed to make the cohort for the complete analysis as large as reliably possible to make a more accurate model.

A previous study performed in the corresponding center aimed at evaluating whether a 6-month interval, as suggested in the European Association for the Study of the Liver guideline, is sufficient to expect HCA regression to <5 cm and showed that the cut-off point for the assessment of regression could be prolonged to at least 12 months (12,13). Time-to-event analysis showed that HCA with larger baseline diameter take considerably longer to regress, as was seen in the current prediction model. The previous study did not find a statistically significant difference between HCA subtype and median time to regression, whereas in the current study, HCA subtype is included in the prediction model. This difference might well be explained by the larger population in the current study (180 vs 118) or the statistical analysis (Kaplan-Meier vs multivariable Cox regression).

This study is subject to a few limitations. Patients were included between 2000 and 2017, a period in which the quality of imaging improved considerably and management might have changed. However, no correlation was found between the year of diagnosis and whether an intervention was performed. Second, interval censoring may occur because imaging provides measurements at a set time point. Third, we used 2-dimensional measurements to assess tumor regression. We are well aware that 2-dimensional measurements do not represent a volume decrease; however, the long-term follow-up indicates a reliable outcome of these measurements. A fourth limitation might lie in the fact that all data were collected retrospectively, and we had missing data for BMI and HCA subtype. Multiple imputation was used to account for the missing data. Finally, the last and most important limitation is the lack of external validation in an independent study population. Although internal validation with bootstrapping techniques suggested a good model fit with minimal overfitting, external validation is preferred before implicating the model in the management of HCA. For external validation, it is custom to have a dataset of at least 100 events, and because HCA is a rare tumor that used to be mostly resected, it will take more time before such a dataset for external validation is available.

In conclusion, it was demonstrated in a large clinical series that a multivariable model comprising of the 3 easily accessible variables, HCA diameter at diagnosis, T0T1 regression-over-time, and HCA subtype, could be helpful to assess the chance of HCA regression in female patients with non-b-catenin–mutated HCA >5 cm at diagnosis that still exceed 5 cm at first follow-up imaging if they adhere to life style adaptations, including cessation of OAC. The model may be of help to clinicians in making a well-informed management decision, reserving invasive treatment only for those patients with a high risk of complications and a low chance of HCA regression to <5 cm. The model still requires external validation in an independent study population. Other investigators are invited to share their data to further improve the risk estimations of the current model.

REFERENCES

1. Edmondson HA, Reynolds TB, Henderson B, et al. Regression of liver cell adenomas associated with oral contraceptives. Ann InternMed 1977;86:180–2.
2. Benton S, Belghiti J. Will weight loss become a future treatment of hepatocellular adenoma in obese patients? Liver Int 2015;35:2228–32.
3. Bunchorntavakul C, Bahirwani R, Drazek D, et al. Clinical features and natural history of hepatocellular adenomas: The impact of obesity. Aliment Pharmacol Ther 2011;34:664–74.
4. Nault JC, Bioulac-Sage P, Zucman-Rossi J. Hepatocellular benign tumors-from molecular classification to personalized clinical care. Gastroenterology 2013;144:888–902.
5. Nault JC, Couchy G, Balabaud C, et al. Molecular classification of hepatocellular adenoma associates with risk factors, bleeding, and malignant transformation. Gastroenterology 2017;152:880–94.e6.
6. van Aalten SM, Thomeer MG, Terkivatan T, et al. Hepatocellular adenomas: Correlation of MR imaging findings with pathologic subtype classification. Radiology 2011;261:172–81.
7. Ronot M, Bahrami S, Calderaro J, et al. Hepatocellular adenomas: Accuracy of magnetic resonance imaging and liver biopsy in subtype classification. Hepatology 2011;53:1182–91.
8. Bieze M, Phoa SS, Verheij J, et al. Risk factors for bleeding in hepatocellular adenoma. Br J Surg 2014;101:847–55.
9. Zucman-Rossi J, Jeannot E, Nhieu JT, et al. Genotype-phenotype correlation in hepatocellular adenoma: New classification and relationship with HCC. Hepatology 2006;43:515–24.
10. Stoot JH, Coelen RJ, De Jong MC, et al. Malignant transformation of hepatocellular adenomas into hepatocellular carcinomas: A systematic review including more than 1600 adenoma cases. HPB (Oxford) 2010;12:509–22.
11. Dejong SM, de Man RA, IJzermans JN, et al. Systematic review of haemorrhage and rupture of hepatocellular adenomas. Br J Surg 2012;99: 911–6.
12. European Association for the Study of the Liver. EASL Clinical Practice Guidelines on the management of benign liver tumours. J Hepatol 2016;65(2):386–98.
13. Terkivatan AJ, Broker MEE, Thomeer MGJ, et al. Retrospective study on timing of resection of hepatocellular adenoma. Br J Surg 2017;104: 1695–703.
14. Laumonier H, Bioulac-Sage P, Laurent C, et al. Hepatocellular adenomas: Magnetic resonance imaging features as a function of molecular pathological classification. Hepatology 2008;48:808–18.
15. Bioulac-Sage P, Cubel G, Taouji S, et al. Immuno-histochemical markers on needle biopsies are helpful for the diagnosis of focal nodular hyperplasia and hepatocellular adenoma subtypes. Am J Surg Pathol 2012;36:1691–9.
16. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation and Updating. New York, NY: Springer, 2009.
17. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. BMJ 2015;350:g7594.

18. de Meijer VE, Kalish BT, Puder M, et al. Systematic review and meta-analysis of steatosis as a risk factor in major hepatic resection. Br J Surg 2010;97:1331–9.
19. Dokmak S, Fteriche FS, Borscheid R, et al. 2012 liver resections in the 21st century: We are far from zero mortality. HPB (Oxford) 2013;15: 908–15.
20. Klompenhouwer AJ, Sprengers D, Willemssen FE, et al. Evidence of good prognosis of hepatocellular adenoma in post-menopausal women. J Hepatol 2016;65:1163–70.

Chapter 6

Development of a stratification tool to identify intraductal papillary mucinous neoplasms at low risk of progression

Kasper A. Overbeek, Maaike Alblas, Valerie Gausman, Pujan Kandel, Adam B. Schweber, Christian Brooks, Priscilla A. Van Riet, Michael B. Wallace, Tamas A. Gonda, Djuna L. Cahen, Marco J. Bruno

Alimentary Pharmacology & Therapeutics. 2019

SUMMARY

Background: Because most pancreatic intraductal papillary mucinous neoplasms (IPMNs) will never become malignant, currently advocated long-term surveillance is low-yield for most individuals.

Aim: To develop a score chart identifying IPMNs at lowest risk of developing worrisome features or high-risk stigmata.

Methods: We combined prospectively maintained pancreatic cyst surveillance data-bases of three academic institutions. Patients were included if they had a presumed side-branch IPMN, without worrisome features or high-risk stigmata at baseline (as de-fined by the 2012 international Fukuoka guidelines), and were followed ≥ 12 months. The endpoint was development of one or more worrisome features or high-risk stigmata during follow-up. We created a multivariable prediction model using Cox- proportional logistic regression analysis and performed an internal-external validation.

Results: 875 patients were included. After a mean follow-up of 50 months (range 12-157), 116 (13%) patients developed worrisome features or high-risk stigmata. The final model included cyst size (HR 1.12, 95% CI 1.09-1.15), cyst multifocality (HR 1.49, 95% CI 1.01-2.18), ever having smoked (HR 1.40, 95% CI 0.95-2.04), history of acute pancreatitis (HR 2.07, 95% CI 1.21-3.55), and history of extrapancreatic malignancy (HR 1.34, 95% CI 0.91-1.97). After validation, the model had good discriminative ability (C-statistic 0.72 in the Mayo cohort, 0.71 in the Columbia cohort, 0.64 in the Erasmus cohort).

Conclusion: In presumed side branch IPMNs without worrisome features or high-risk stigmata at baseline, the Dutch-American Risk stratification Tool (DART-1) success-fully identifies pancreatic lesions at low risk of developing worrisome features or high-risk stigmata.

INTRODUCTION

Pancreatic cystic lesions are a common, often incidental finding. Recent large studies using magnetic resonance cholangiopancreatography revealed a remarkably high prevalence in the general population[1, 2] of up to 49% and even up to 60% for persons over 70 years.[2] Many of these lesions are neoplastic mucinous cysts, a subgroup with a varying risk of malignant progression, depending on pathological subtype and extent of pancreatic duct involvement.

Of all neoplastic cysts, side branch intraductal papillary mucinous neoplasms (SB-IPMN) are the most common and deemed to bear the lowest risk of harboring malignancy or progressing to malignancy. Risk estimations were initially based on small, retrospective series, evaluating mainly resected SB-IPMN in tertiary referral centers.[3-6] They reported a risk of invasive carcinoma ranging from 11%6 to 29%.[3] However, several recent studies indicate a much lower risk for incidentally found SB-IPMN. In 2015, a meta-analysis was published including 2177 patients under surveillance for SB-IPMN, of which only 82 (3.7%) developed a pancreatic malignancy.[7] Since then, several additional studies, each including at least 300 patients with at least five years of follow-up, reported a pancreatic cancer risk of only 0-1.6% for small asymptomatic cysts.[2, 8-12] However, all these studies were retrospective and the actual, long-term risk is yet to be determined by large and prospective studies.

Pending definite answers, the European,[13] AGA,[14] ACG,[15] and international Fukuoka[16] guidelines recommend surveillance with magnetic resonance imaging/magnetic resonance cholangiopancreatography and/or endoscopic ultrasound for all IPMNs, including small unsuspected cysts, in an attempt to detect pancreatic cancer in an early or even premalignant stage. These recommendations pose a considerable burden on patients and health care resources, while the clinical benefit with regard to survival remains to be proven.

There are currently no tools to distinguish IPMNs that do not warrant surveillance, or that are helpful in selecting a tailored and optimal surveillance interval. Previous prediction models have focused on identifying high-risk IPMNs to improve patient selection for surgery.[17-24] Although these models are valuable and necessary, the vast majority of SB-IPMNs do not progress. Therefore, we aimed to develop a prediction model that identifies patients with SB-IPMN at lowest risk of developing worrisome features or high-risk stigmata. Such a stratifying tool is needed to prevent redundant surveillance and reduce the burden for patients and health care systems.

MATERIALS AND METHODS

Study design

We included pancreatic cyst surveillance data from prospectively-maintained databases of three academic institutions, namely the Erasmus University Medical Center, Rotterdam, the Netherlands; Columbia University Medical Center, New York, USA; and the Mayo Clinic Florida, Jacksonville, USA. At the Erasmus UMC, the study was exempt from institutional review board review (MEC-2018-1285). The study received IRB approval at Columbia UMC (AAAO8260(M01Y04)) and at the Mayo Clinic (14-007100). The need for written informed consent was waived by the Erasmus UMC and Columbia UMC. At the Mayo Clinic Florida, verbal informed consent was obtained from each participant before enrollment. The study was performed according to the declaration of Helsinki and the manuscript complies with the statement for the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD).[25]

Participants

The databases contain all consecutive patients under surveillance for a pancreatic cyst since 2004 (Erasmus University Medical Center), 2003 (Columbia University Medical Center), and 2000 (Mayo Clinic Florida). From these databases, we selected patients with a radiologically presumed SB-IPMN who had been followed-up for at least 12 months. A subset of these patients have been described previously.[26] We excluded individuals with one or more worrisome features or high-risk stigmata at baseline, as defined in the 2012 international Fukuoka guidelines[27] (Figure 1).

Endpoint and candidate predictors

The endpoint was defined as the development of one or more worrisome features or high-risk stigmata according to the 2012 International Fukuoka guidelines. Candidate predictors were chosen based on prior publications and medical reasoning. Included in the analysis were age, personal history of diabetes mellitus (defined as having a previous diagnosis in electronic medical records), body mass index, having smoked ever, personal history of acute pancreatitis, personal history of any type of extrapancreatic malignancy, family history of pancreatic ductal adenocarcinoma, multifocality of the cyst, and the diameter of the largest cyst. All variables were assessed at the time of cyst diagnosis.

Statistical analysis

Missing data was imputed using multiple imputation by chained equations (MICE) based on the posterior distributions with five datasets with the MICE package in R software.[28] We used a Cox-proportional logistic regression analysis to develop a multivariable prediction model. A linear relation was the best approximation of the relationship between the endpoint and the continuous predictors. A backward stepwise selection procedure was

performed with Akaike's Information Criterion as stopping rule, to select the model with the highest predictive performance. This final model was presented with hazard ratios, and 95% confidence intervals calculated using a parametric approach, to indicate the individual predictor effects. Because of interrelationships among predictors, in this approach the final model can include predictors that have non-statistically significant individual predictor effects. The Cox-proportional hazard assumption was checked and showed non-significant results, indicating that proportional hazards can be assumed.

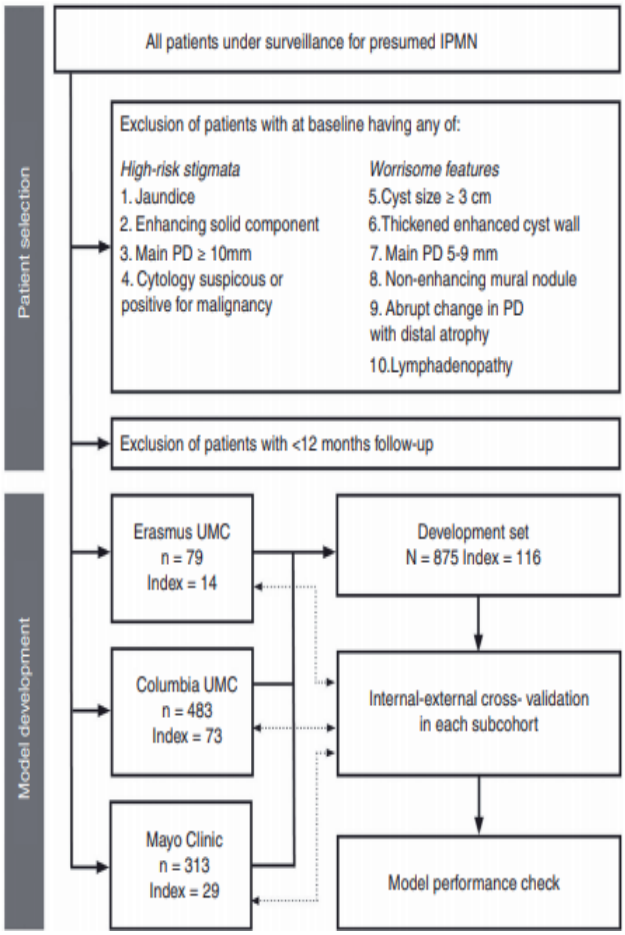

FIGURE 1. Flow-chart of patient selection and model development process

We first performed an internal validation with bootstrap resampling with 500 replications to shrink the model's coefficients to minimize overfitting.[29] Subsequently, we performed an internal-external validation of the final model, in which each subcohort was in turn omitted from the development set and subsequently used as validation set (Figure 1). Model

performance in terms of discriminative ability was described with the Harrell's concordance statistic (C-statistic), which varies between 0.5 (a non-informative model) and 1.0 (a perfect model). The coefficients were used to calculate the probability of developing worrisome features or high-risk stigmata within three years and within five years, which is presented in a score chart. We used SPSS Statistics 22 (IBM Corporation, Armonk, New York, USA) and R Software version 3.3.5 (R foundation for statistical computing, Vienna, Austria) for the statistical analysis.

RESULTS

Participants and clinical outcome

We included 875 patients. The mean age was 66 (SD 11.2) years, 37% (321) were male, 74% (648) Caucasian, and the mean BMI was 27 (SD 4.9). At baseline, multifocal cysts were observed in 335 (38%) patients and the average diameter of the largest cyst was 12 mm (SD 6.4, see table 1 for all baseline characteristics). After a mean follow-up of 50 months (SD 28.5, range 12-157) and a total follow-up of 3,649 person-years, 116 (13.2%) patients developed one or more worrisome features or high-risk stigmata. Table 2 shows the baseline characteristics according to outcome.

In the group who developed a worrisome feature, surgery was performed on 36 (31%) patients. Pathology showed an invasive carcinoma in 3, high grade dysplasia in 6, low or moderate grade dysplasia in 22, a neuroendocrine tumor in 1, and a mucinous cystic neoplasm in 4 patients. In the group without a worrisome feature during follow-up, surgery was performed on 20 (2.6%) patients. Reasons for this included the presence of symptoms (other than jaundice or current pancreatitis), minor growth of a cyst smaller than 3 cm, an increased cyst fluid carcinoembryonic antigen level, a pancreatitis episode in the past (but not at the moment of cyst detection), the patient's wishes, or a combination of these reasons. In these cases, pathology showed only low or moderate grade dysplasia (18) or a mucinous cystic neoplasm (2). Of the non-operated patients, none were diagnosed with pancreatic cancer during follow-up.

Missing data and model specification

None of the patients had missing data for the endpoint, age, cyst multifocality, or initial cyst size. There was ≤5% missing data for smoking behavior (4.5%), personal history of diabetes (0.6%), personal history of acute pancreatitis (2.1%), personal history of extrapancreatic malignancy (1.0%), and family history of pancreatic ductal adenocarcinoma (4.3%). For body mass index, data was missing for 200 (23%) patients.

The model with the best fit included cyst size (HR 1.12, 95% CI 1.09-1.15), cyst multifocality (HR 1.49, 95% CI 1.01-2.18), having smoked ever (HR 1.40, 95% CI 0.95-2.04), history of acute pancreatitis (HR 2.07, 95% CI 1.21-3.55), and history of extrapancreatic malignancy (HR1.34, 95% CI 0.91-1.97). The hazard ratios and 95% CI of each predictive variable in both univariable and multivariable analysis are shown in table 3.

TABLE 1. Baseline patient and cyst characteristics.

	Erasmus UMC (n=79)	Columbia UMC (n=483)	Mayo Clinic Florida (n=313)
Patient characteristics			
Age, mean (SD), y	61 (11.0)	65 (11.9)	68 (9.5)
Male gender	20 (25.3)	197 (40.8)	104 (33.2)
Race			
Caucasian	64 (81.0)	295 (61.1)	289 (92.3)
Asian	2 (2.5)	20 (4.1)	3 (1.0)
Black	4 (5.1)	30 (6.2)	16 (5.1)
Other	4 (5.1)	16 (3.3)	1 (0.3)
Unknown	5 (6.3)	122 (25.3)	4 (1.3)
Diabetes mellitus	10 (12.7)	119 (24.6)	46 (14.7)
Body mass index, mean (SD)	27 (5.6)	27 (5.0)	27 (4.8)
Smoking ever	27 (34.2)	189 (39.1)	126 (40.3)
Alcohol ever	38 (48.1)	198 (41)	136 (43.5)
History of acute pancreatitis	9 (11.3)	48 (9.9)	13 (4.2)
History of extrapancreatic malignancy	12 (15.2)	195 (40.4)	84 (26.8)
Family history of PDAC	10 (12.7)	50 (10.4)	30 (9.6)
Cyst characteristics			
Location dominant cyst			
Head	52 (65.8)	188 (38.9)	141 (45.0)
Body	22 (27.8)	188 (38.9)	103 (32.9)
Tail	4 (5.1)	106 (21.9)	68 (21.7)
Multifocality	43 (54.4)	188 (38.9)	104 (33.2)
Largest diameter, mean (SD), mm	13 (6.6)	11.5 (6.5)	12 (6.1)

Values presented as n (%) unless otherwise indicated; SD, standard deviation; PDAC, pancreatic ductal adenocarcinoma;

Model performance

Bootstrap resampling showed limited optimism in the C-statistic of 0.02. In the internal-external validation, model performance varied between the three subcohorts. The model showed the best discriminative ability in the cohorts of Mayo Clinic Florida (C-statistic 0.72, 95% CI 0.61-0.84) and Columbia UMC (C-statistic 0.71, 95% CI 0.66-0.80). The performance within the Erasmus UMC cohort was 0.64 (95% CI 0.57-0.88).

Score chart and example

The Dutch-American Risk stratification Tool (DART-1) visualises the estimated 3-year and 5-year risk of developing one or more worrisome features or high-risk stigmata for all possible predictor combinations (Figures 2A and 2B). A web-based application has been developed and is available at https://rtools.mayo.edu/DART/ (Figure 3). When using the DART-1, a patient with a unifocal cyst smaller than 1 cm, without a history of acute pancreatitis, extrapancreatic malignancy or smoking, has an estimated 3-year risk of ≤2% and 5-year risk of ≤5% to develop one or more worrisome features or high-risk stigmata.

TABLE 2. Patient and cyst characteristics separated on study endpoint.

	Total (N = 875)	No development of WF or HRS (n = 759)	Development of WF or HRS (n = 116)
Center			
Erasmus UMC	79 (9.0)	65 (8.6)	14 (12.1)
Columbia UMC	483 (55.2)	410 (54.0)	73 (62.9)
Mayo Clinic	313 (35.8)	284 (37.4)	29 (25.0)
Patient characteristics			
Age, mean (SD), y	66 (11.2)	65 (10.9)	67 (12.8)
Male gender	321 (36.7)	271 (35.7)	50 (43.1)
Race			
Caucasian	648 (74.1)	568 (74.8)	80 (69.0)
Asian	25 (2.9)	22 (2.9)	3 (2.6)
Black	50 (5.7)	41 (5.4)	9 (7.8)
Other	21 (2.3)	18 (2.4)	3 (2.6)
Unknown	131 (15.0)	110 (14.5)	21 (18.1)
Diabetes mellitus	175 (20.0)	148 (19.5)	27 (23.3)
Body mass index, mean (SD)	27 (4.9)	27 (4.9)	27 (5.3)
Smoking ever	342 (39.1)	288 (37.9)	54 (46.6)
Alcohol ever	372 (42.5)	319 (42.0)	53 (45.7)
History of acute pancreatitis	70 (8.0)	54 (7.1)	16 (13.8)
History of extrapancreatic malignancy	291 (33.3)	246 (32.4)	45 (38.8)
Family history of PDAC	90 (10.3)	80 (10.5)	10 (8.6)
Cyst characteristics			
Location dominant cyst			
Head	381 (43.5)	329 (43.3)	52 (44.8)
Body	313 (35.8)	274 (36.1)	39 (33.6)
Tail	178 (20.3)	153 (20.2)	25 (21.6)
Multifocality	335 (38.3)	280 (36.9)	55 (47.4)
Largest diameter, mean (SD), mm	12 (6.4)	11 (6.0)	17 (6.7)

Values presented as n (%) unless otherwise indicated; WF, worrisome feature; HRS, high-risk stigmata; SD, standard deviation; PDAC, pancreatic ductal adenocarcinoma.

TABLE 3. Candidate predictors with associated hazard ratios.

	Univariable		Final multivariable model	
Predictor	**HR**	**95% CI**	**HR**	**95% CI**
Age	1.01	0.99-1.03	NA	NA
Body mass index	1.01	0.96-1.05	NA	NA
Smoking, ever	1.42	0.98-2.05	1.40	0.95-2.04
History of diabetes mellitus	1.37	0.89-2.12	NA	NA
History of acute pancreatitis	1.76	1.04-2.99	2.07	1.21-3.55
History of extrapancreatic malignancy	1.21	0.83-1.76	1.34	0.91-1.97
Cyst multifocality	1.65	1.14-2.41	1.49	1.01-2.18
Largest cyst diameter, per mm	1.12	1.09-1.15	1.12	1.09-1.15

HR, Hazard Ratio; CI, Confidence Interval; NA, Not applicable, was not included in the model with the best fit;

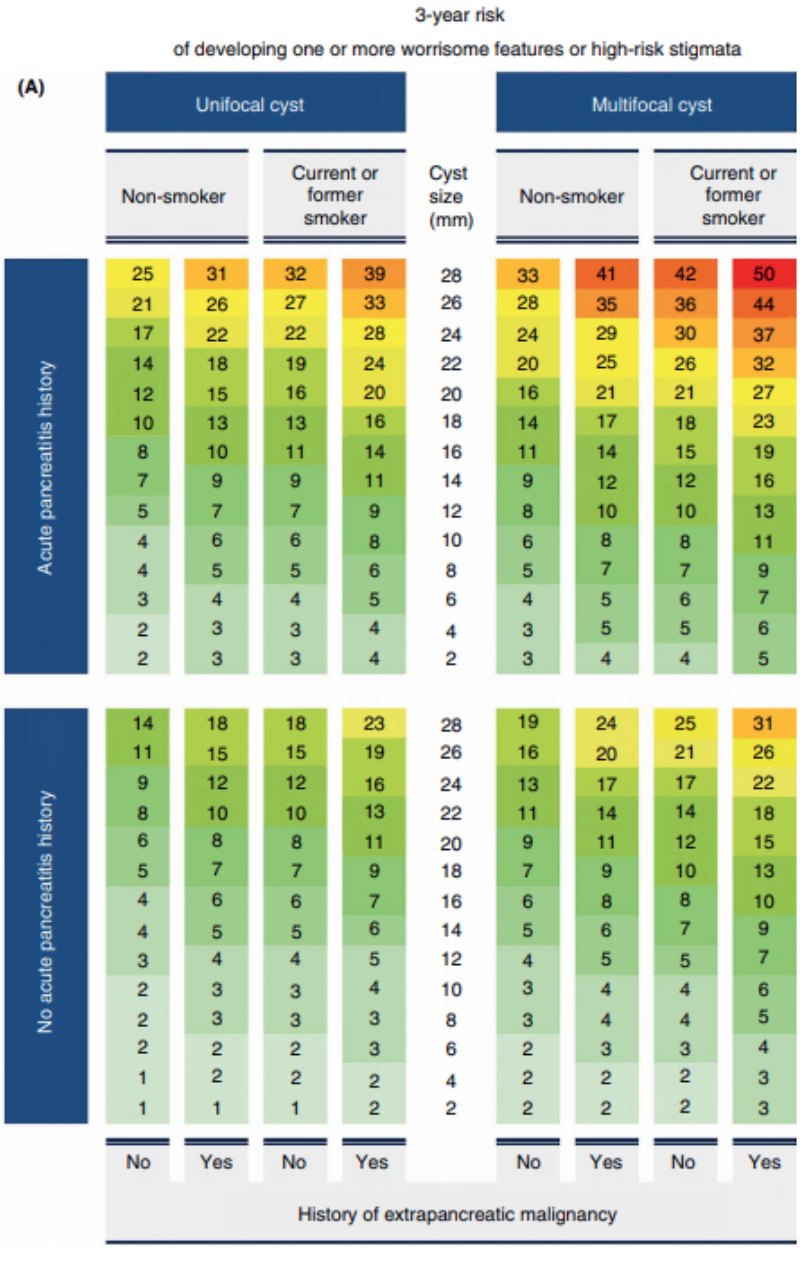

FIGURE 2A. The Dutch-American Risk stratification Tool (DART-1) to identify SB-IPMN at low probability (%) of developing one or more worrisome features or high-risk stigmata within 3 years.

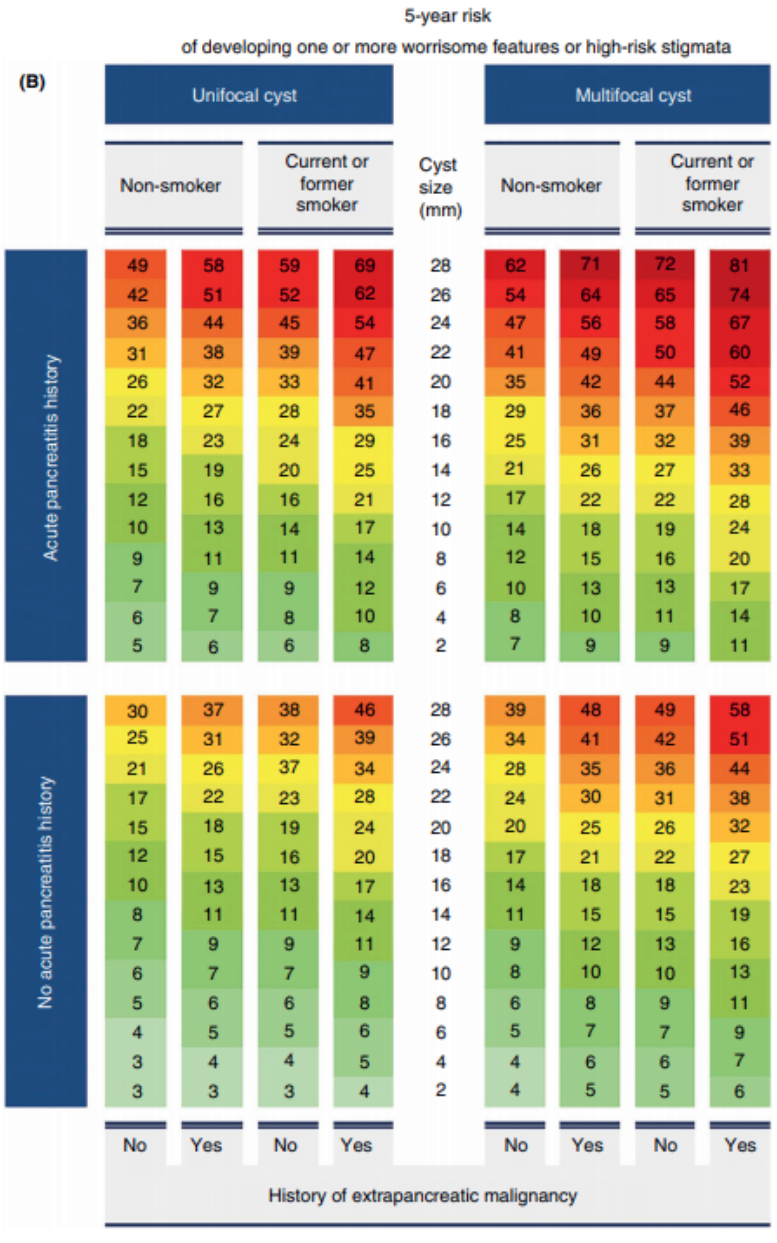

FIGURE 2B. The Dutch-American Risk stratification Tool (DART-1) to identify SB-IPMN at low probability (%) of developing one or more worrisome features or high-risk stigmata within 5 years.

DISCUSSION

In this international multicenter study, we describe the development of DART-1, the first version of a prediction model that does not focus on identifying IPMNs at high risk of malignancy, but on those at low risk instead. It is based on patient and cyst characteristics that can be assessed at the time of diagnosis, and predicts the 3-year and 5-year risk of developing worrisome features or high-risk stigmata as defined by the 2012 international Fukuoka guidelines. Such a model is important, as pancreatic cysts are diagnosed with increasing frequency and yearly imaging is generally recommended, even though the majority of lesions are at low risk of malignant progression. By using a stratifying tool, clinicians can make evidence-based risk estimations for progression in individual patients and identify those at lowest risk. The ultimate goal would be to decrease the burden of surveillance on patients, but also on health care resources by either optimizing surveillance intervals or, in selected cases, discontinue surveillance.

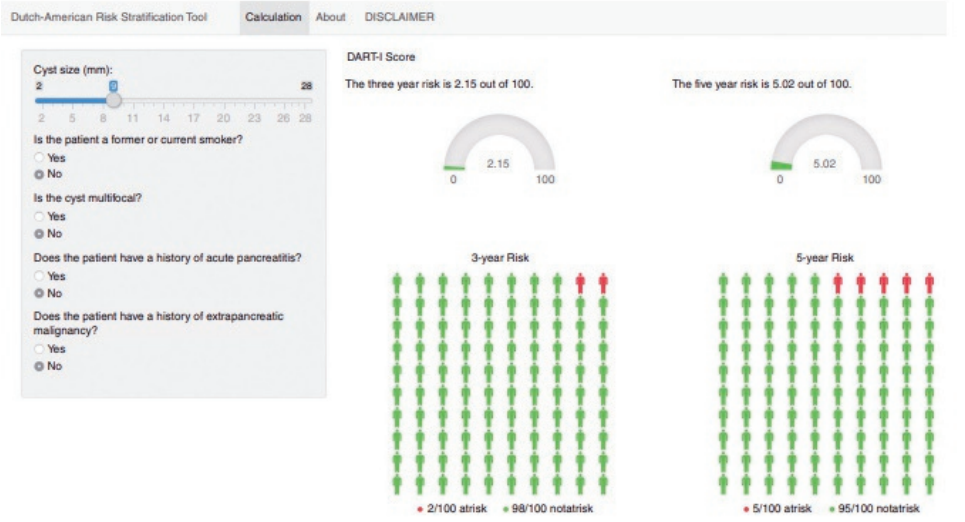

FIGURE 3. The web-based application of the Dutch-American Risk stratification Tool (DART-1) with an example patient with low probability of developing one or more worrisome features or high-risk stigmata. The application can be found at https://rtools.mayo.edu/DART/.

In our cohort, multivariable analysis resulted in five predictors for progression: cyst size, cyst multifocality, having smoked ever, history of acute pancreatitis, and history of extrapancreatic malignancy. Cyst size being an independent predictor of progression comes as no surprise, given that a size of 3 centimeters or greater is defined as a worrisome feature[27] and therefore incorporated in our composite endpoint. However, it has been shown in other cohorts that initial cyst size is a predictor of cyst growth,[30-32] development of other worrisome features,[31, 33] and malignancy.[12] The predictive value of cyst multifocality has been described less often,

but is not a new finding. Crippa et al. followed 144 patients with SB-IPMN for 5 years, and found that an increase in the number of lesions was associated with the development of worrisome features or high-risk stigmata (OR 6, 95% CI 1.7-20.8).[33] It was also identified as predictor in an earlier analysis of a subset of our cohort.[26] A history of smoking and of acute pancreatitis are well-established risk factors for pancreatic cancer,[34-37] but not for the development of worrisome features in IPMN. Some studies suggest smoking accelerates progression of IPMN, and that it predicts invasive IPMN or concomitant pancreatic cancer in resected IPMN, but results are conflicting.[38-41]

A history of extrapancreatic malignancy has not been described as a predictor for progression in other cohorts. In the previous analysis of a subset of our cohort, a history of any extrapancreatic malignancy was not an independent predictor, but a history of prostate cancer was. This difference is most likely attributable to the difference in sample size. Retrospective studies have reported an increased incidence of extrapancreatic malignancies in patients with IPMN, but prospective studies were unable to confirm this.[8, 42] Crippa and colleagues did not find an association between extrapancreatic tumors and the development of worrisome features,[33] but because their cohort consisted of 144 patients of which only 26 developed worrisome features or high-risk stigmata, this may be due to a lack of power. The predictive value of a history of an extrapancreatic malignancy on progression of IPMN has to be confirmed by studies in other cohorts.

Having a history of diabetes was predictive in the univariable analysis but did not contribute significantly to the multivariable model and was therefore omitted from DART-1. The association between diabetes and pancreatic cancer is well-known,[43-46] but the association with IPMN is less established. Some studies have reported an increased risk for patients with diabetes to develop IPMN,[1, 47] but in another large population-based study, this association disappeared after correcting for age and body mass index.2 Morales-Oyarvide et al. showed that in patients with resected IPMN, preoperative diabetes is associated with high-grade dysplasia and invasive carcinoma,[48] suggesting diabetes has a proliferative effect on the cyst. However, to our knowledge, there have been no studies that demonstrate that diabetes is associated with the development of worrisome features and high-risk stigmata. Although diabetes did not contribute to the predictive ability of the model in our cohort, it should be included in validation studies and future updates of DART-1, to further establish its value.

We encountered some minor differences between the subcohorts, the most noticeable being a higher prevalence of diabetes and personal history of extrapancreatic malignancy in the Columbia cohort, and more multifocal cysts in the Erasmus cohort. However, any meaningful differences between the subcohorts were ruled out by the internal-external validation. In this type of validation, each subcohort is in turn left out from the development set and used as a validation set. The final model is then based on all available data. Such an internal-

external cross-validation can be used to demonstrate external validity of a prediction model, with the additional advantage that sample size is retained.[49] DART-1 performed similarly in the total cohort before validation (apparent performance), the Columbia cohort, and the Mayo cohort. The slight decrease in performance within the Erasmus cohort was expected and is attributable to this cohort's smaller sample size.

DART-1 shows promise, but should be interpreted with some caution. Foremost, prediction models are developed to augment, and not replace clinical judgment, and the given risks are estimates that therefore hold some extent of uncertainty. Also, it is crucial that DART-1 is validated in other cohorts before it is implemented in clinical care. We expect DART-1 will be highly generalisable because our development set encompasses three centers, each located in a different geographical region, and each collecting patient data in slightly different time periods. Also, our cohort consists of patients without complex cysts, and is therefore likely to be comparable to the patient population in the primary or secondary care setting. Additionally, we observed limited optimism in the C-statistic and, therefore, a good external performance is likely.

The main limitation of this prediction model is that it uses a composite, surrogate endpoint. Ideally it would predict development of malignancy. However, given the low cancer risk of SB-IPMNs, it would require extremely large cohorts to reach adequate numbers for statistical modeling. Although we collected one of the largest low-risk SB-IPMN cohorts, it did not yield enough pancreatic cancer cases for this purpose, and we are unable to make predictions on the development of malignancy. However, the ultimate objective of DART-1 is not to identify high-risk IPMNs, but those unlikely to develop into malignancy. Although it has been shown that worrisome features and high-risk stigmata accurately stratify for malignancy risk,[50] it is also known that a substantial number of IPMNs with a worrisome feature do not harbor high-grade dysplasia or invasive carcinoma,[21-23] which is supported by our own results. IPMNs without worrisome features harbor an even lower risk of developing pancreatic cancer, which strengthens the usefulness of DART-1 as a negative prediction tool that can be used to identify those SB-IPMNs that require less intense surveillance.

A second limitation is that we have based our endpoint on the 2012 international Fukuoka guidelines,[27] whereas these were revised in 2017.[16] Similar to the European guidelines,[13] the updated version includes elevated serum carbohydrate antigen 19-9 levels as a worrisome feature, as well as cyst growth. In our cohorts, serum carbohydrate antigen 19-9 levels and exact cyst growth were not routinely determined and recorded in the past, because they were under surveillance long before guidelines stressed the importance of these parameters. Because previous studies have shown that cysts not necessarily display a linear growth pattern[31, 32] and that there is a variability in size measurement between imaging modalities[51] and between observers[52], it was not possible to reliably assess growth rate

retrospectively. Therefore, we could not use the updated guidelines, and fast-growing IPMNs that did not reach 3 centimeters during the follow-up period, may have been misidentified as non-progressors. Now that serum carbohydrate antigen 19-9 levels and growth rates are routinely determined as per guidelines, it will be possible to include these variables as part of the study endpoint or as predictor in future updates. Another aspect that could not be completely ruled out, is if our dataset contained a bias by right censoring. However, the predictors in the model did not show an association with follow-up time, limiting the possible influence of this type of bias.

An issue of much debate is whether the risk of malignancy increases over time. Some recent studies have shown that even small SB-IPMN may evolve into malignancy after 5 or 10 years,[11, 12, 33] and that a stable cyst size for 5 years does not preclude future growth.[53] Because our study population has a mean follow-up of 50 months, we are not yet able to determine these long-term risks. At this point in time, this precludes us from stopping surveillance altogether, based on DART-1. Therefore, we do not advocate a complete stop of surveillance, but suggest a reduction of surveillance frequency for the lowest risk SB-IPMNs, the ideal cut-off for which requires further calculation in external cohorts. It is essential to update DART-1 based on long-term, prospective data. Additional predictors should be explored, such as diabetes, glycated haemoglobin or serum fasting glucose, serum carbohydrate antigen 19-9 level, or other promising biomarkers. It may also be of interest to objectify smoking exposure, that is, using pack years as a predictor rather than a history of smoking. It is also conceivable that current smokers are at higher risk than former smokers. Cyst growth may also be a strong predictor, but including this will render the model unfit for use at the time of cyst diagnosis.

In conclusion, we have developed a prediction model that does not focus on detecting high-risk IPMNs, but identifies IPMNs at lowest risk of developing worrisome features or high-risk stigmata instead, by combining variables readily available at the time of cyst diagnosis. Even though DART-1 is the first version of this type of prediction model, it had a good performance in an internal-external validation, and high generalisability to other cohorts is expected. After DART-1 is externally validated by others, it can be used to explore varying surveillance strategies using looser follow-up policies for IPMNs at lowest risk. This very novel approach of stratifying IPMNs has the potential to protect patients with low-risk IPMNs from redundant medical interventions, and to reduce costs and the burden for the health care system.

References

1. Mizuno S, Isayama H, Nakai Y, et al. Prevalence of Pancreatic Cystic Lesions Is Associated With Diabetes Mellitus and Obesity: An Analysis of 5296 Individuals Who Underwent a Preventive Medical Examination. Pancreas 2017;46:801-805.
2. Kromrey ML, Bulow R, Hubner J, et al. Prospective study on the incidence, prevalence and 5-year pancreatic-related mortality of pancreatic cysts in a population-based study. Gut 2018;67:138-145.
3. Levy P, Jouannaud V, O'Toole D, et al. Natural history of intraductal papillary mucinous tumors of the pancreas: actuarial risk of malignancy. Clin Gastroenterol Hepatol 2006;4:460-8.
4. Rodriguez JR, Salvia R, Crippa S, et al. Branch-duct intraductal papillary mucinous neoplasms: observations in 145 patients who underwent resection. Gastroenterology 2007;133:72-9; quiz 309-10.
5. Schnelldorfer T, Sarr MG, Nagorney DM, et al. Experience with 208 resections for intraductal papillary mucinous neoplasm of the pancreas. Arch Surg 2008;143:639-46; discussion 646.
6. Crippa S, Fernandez-Del Castillo C, Salvia R, et al. Mucin-producing neoplasms of the pancreas: an analysis of distinguishing clinical and epidemiologic characteristics. Clin Gastroenterol Hepatol 2010;8:213-9.
7. Crippa S, Capurso G, Camma C, et al. Risk of pancreatic malignancy and mortality in branch-duct IPMNs undergoing surveillance: A systematic review and meta-analysis. Dig Liver Dis 2016;48:473-9.
8. Malleo G, Marchegiani G, Borin A, et al. Observational study of the incidence of pancreatic and extrapancreatic malignancies during surveillance of patients with branch-duct intraductal papillary mucinous neoplasm. Ann Surg 2015;261:984-90.
9. Kwong WT, Hunt GC, Fehmi SM, et al. Low Rates of Malignancy and Mortality in Asymptomatic Patients With Suspected Neoplastic Pancreatic Cysts Beyond 5 Years of Surveillance. Clin Gastroenterol Hepatol 2016;14:865-871.
10. Munigala S, Gelrud A, Agarwal B. Risk of pancreatic cancer in patients with pancreatic cyst. Gastrointest Endosc 2016;84:81-6.
11. Del Chiaro M, Ateeb Z, Hansson MR, et al. Survival Analysis and Risk for Progression of Intraductal Papillary Mucinous Neoplasia of the Pancreas (IPMN) Under Surveillance: A Single-Institution Experience. Ann Surg Oncol 2017;24:1120-1126.
12. Pergolini I, Sahora K, Ferrone CR, et al. Long-term Risk of Pancreatic Malignancy in Patients With Branch Duct Intraductal Papillary Mucinous Neoplasm in a Referral Center. Gastroenterology 2017;153:1284-1294.e1.
13. European evidence-based guidelines on pancreatic cystic neoplasms. Gut 2018;67:789-804.
14. Vege SS, Ziring B, Jain R, et al. American gastroenterological association institute guideline on the diagnosis and management of asymptomatic neoplastic pancreatic cysts. Gastroenterology 2015;148:819-22.
15. Elta GH, Enestvedt BK, Sauer BG, et al. ACG Clinical Guideline: Diagnosis and Management of Pancreatic Cysts. Am J Gastroenterol 2018;113:464-79.

16. Tanaka M, Fernandez-Del Castillo C, Kamisawa T, et al. Revisions of international consensus Fukuoka guidelines for the management of IPMN of the pancreas. Pancreatology 2017;17:738-753.
17. Buscaglia JM, Giday SA, Kantsevoy SV, et al. Patient- and cyst-related factors for improved prediction of malignancy within cystic lesions of the pancreas. Pancreatology 2009;9:631-8.
18. Matthaei H, Wylie D, Lloyd MB, et al. miRNA biomarkers in cyst fluid augment the diagnosis and management of pancreatic cysts. Clin Cancer Res 2012;18:4713-24.
19. Hijioka S, Shimizu Y, Mizuno N, et al. Can long-term follow-up strategies be determined using a nomogram-based prediction model of malignancy among intraductal papillary mucinous neoplasms of the pancreas? Pancreas 2014;43:367-72.
20. Shimizu Y, Yamaue H, Maguchi H, et al. Validation of a nomogram for predicting the probability of carcinoma in patients with intraductal papillary mucinous neoplasm in 180 pancreatic resection patients at 3 high-volume centers. Pancreas 2015;44:459-64.
21. Attiyeh MA, Fernandez-Del Castillo C, Al Efishat M, et al. Development and Validation of a Multi-Institutional Preoperative Nomogram for Predicting Grade of Dysplasia in Intraductal Papillary Mucinous Neoplasms (IPMNs) of the Pancreas: A Report from The Pancreatic Surgery Consortium. Ann Surg 2016.
22. Jang JY, Park T, Lee S, et al. Proposed Nomogram Predicting the Individual Risk of Malignancy in the Patients With Branch Duct Type Intraductal Papillary Mucinous Neoplasms of the Pancreas. Ann Surg 2016.
23. Suzuki Y, Nakazato T, Yokoyama M, et al. Development and Potential Utility of a New Scoring Formula for Prediction of Malignant Intraductal Papillary Mucinous Neoplasm of the Pancreas. Pancreas 2016;45:1227-32.
24. Al Efishat MA, Attiyeh MF, Eaton AA, et al. Multi-Institutional Validation Study of Pancreatic Cyst Fluid Protein Analysis for Prediction of High-Risk Intraductal Papillary Mucinous Neoplasms of the Pancreas. Ann Surg 2017.
25. Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. Br J Surg 2015;102:148-58.
26. Gausman V, Kandel P, Van Riet PA, et al. Predictors of Progression Among Low-Risk Intraductal Papillary Mucinous Neoplasms in a Multicenter Surveillance Cohort. Pancreas 2018;47:471-476.
27. Tanaka M, Fernandez-del Castillo C, Adsay V, et al. International consensus guidelines 2012 for the management of IPMN and MCN of the pancreas. Pancreatology 2012;12:183-97.
28. Van Buuren S. Flexible imputation of missing data. Boca Raton, FL: Chapman & Hall/CRC, 2012.
29. Steyerberg EW. Clinical Prediction Models. New York: Springer-Verlag, 2009.
30. Kim GE, Shin SS, Kim JW, et al. Incidental, Small (< 3 cm), Unilocular, Pancreatic Cysts: Factors That Predict Lesion Progression during Imaging Surveillance. Korean J Radiol 2017;18:915-925.
31. Han Y, Lee H, Kang JS, et al. Progression of Pancreatic Branch Duct Intraductal Papillary Mucinous Neoplasm Associates With Cyst Size. Gastroenterology 2018;154:576-584.
32. Kayal M, Luk L, Hecht EM, et al. Long-Term Surveillance and Timeline of Progression of Presumed Low-Risk Intraductal Papillary Mucinous Neoplasms. AJR Am J Roentgenol 2017;209:320-326.

33. Crippa S, Pezzilli R, Bissolati M, et al. Active Surveillance Beyond 5 Years Is Required for Presumed Branch-Duct Intraductal Papillary Mucinous Neoplasms Undergoing Non-Operative Management. Am J Gastroenterol 2017;112:1153-1161.
34. Yu A, Woo SM, Joo J, et al. Development and Validation of a Prediction Model to Estimate Individual Risk of Pancreatic Cancer. PLoS One 2016;11:e0146473.
35. Risch HA, Yu H, Lu L, et al. Detectable Symptomatology Preceding the Diagnosis of Pancreatic Cancer and Absolute Risk of Pancreatic Cancer Diagnosis. Am J Epidemiol 2015;182:26-34.
36. Bosetti C, Lucenteforte E, Silverman DT, et al. Cigarette smoking and pancreatic cancer: an analysis from the International Pancreatic Cancer Case-Control Consortium (Panc4). Ann Oncol 2012;23:1880-8.
37. Kirkegard J, Cronin-Fenton D, Heide-Jorgensen U, et al. Acute Pancreatitis and Pancreatic Cancer Risk: A Nationwide Matched-Cohort Study in Denmark. Gastroenterology 2018;154:1729-1736.
38. Carr RA, Roch AM, Shaffer K, et al. Smoking and IPMN malignant progression. Am J Surg 2017;213:494-497.
39. Kamata K, Takenaka M, Nakai A, et al. Association between the Risk Factors for Pancreatic Ductal Adenocarcinoma and Those for Malignant Intraductal Papillary Mucinous Neoplasm. Oncology 2017;93 Suppl 1:102-106.
40. Rezaee N, Khalifian S, Cameron JL, et al. Smoking is not associated with severe dysplasia or invasive carcinoma in resected intraductal papillary mucinous neoplasms. J Gastrointest Surg 2015;19:656-65.
41. Nakagawa T, Masuda A, Toyama H, et al. Smoking Status and the Incidence of Pancreatic Cancer Concomitant With Intraductal Papillary Mucinous Neoplasm. Pancreas 2017;46:582-588.
42. Pugliese L, Keskin M, Maisonneuve P, et al. Increased incidence of extrapancreatic neoplasms in patients with IPMN: Fact or fiction? A critical systematic review. Pancreatology 2015;15:209-16.
43. Everhart J, Wright D. Diabetes mellitus as a risk factor for pancreatic cancer. A meta-analysis. Jama 1995;273:1605-9.
44. Huxley R, Ansary-Moghaddam A, Berrington de Gonzalez A, et al. Type-II diabetes and pancreatic cancer: a meta-analysis of 36 studies. Br J Cancer 2005;92:2076-83.
45. Ben Q, Xu M, Ning X, et al. Diabetes mellitus and risk of pancreatic cancer: A meta-analysis of cohort studies. Eur J Cancer 2011;47:1928-1937.
46. Batabyal P, Vander Hoorn S, Christophi C, et al. Association of diabetes mellitus and pancreatic adenocarcinoma: a meta-analysis of 88 studies. Ann Surg Oncol 2014;21:2453-62.
47. Capurso G, Boccia S, Salvia R, et al. Risk factors for intraductal papillary mucinous neoplasm (IPMN) of the pancreas: a multicentre case-control study. Am J Gastroenterol 2013;108:1003-9.
48. Morales-Oyarvide V, Mino-Kenudson M, Ferrone CR, et al. Diabetes mellitus in intraductal papillary mucinous neoplasm of the pancreas is associated with high-grade dysplasia and invasive carcinoma. Pancreatology 2017.
49. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 2016;69:245-7.

50. Mukewar S, de Pretis N, Aryal-Khanal A, et al. Fukuoka criteria accurately predict risk for adverse outcomes during follow-up of pancreatic cysts presumed to be intraductal papillary mucinous neoplasms. Gut 2017;66:1811-1817.
51. Maimone S, Agrawal D, Pollack MJ, et al. Variability in measurements of pancreatic cyst size among EUS, CT, and magnetic resonance imaging modalities. Gastrointest Endosc 2010;71:945-50.
52. Dunn DP, Brook OR, Brook A, et al. Measurement of pancreatic cystic lesions on magnetic resonance imaging: efficacy of standards in reducing inter-observer variability. Abdom Radiol (NY) 2016;41:500-7.
53. Lawrence SA, Attiyeh MA, Seier K, et al. Should Patients With Cystic Lesions of the Pancreas Undergo Long-term Radiographic Surveillance?: Results of 3024 Patients Evaluated at a Single Institution. Ann Surg 2017;266:536-544.

Chapter 7

Risk of a primary keratinocyte carcinoma in patients with actinic keratosis: a prognostic model

S. Tokez, M. Alblas, T. Nijsten, L.M. Pardo Cortes, H.F. Lingsma, M. Wakkee

British Journal of Dermatology. 2019

SUMMARY

Background: Patients with actinic keratoses (AKs) are at increased risk for developing keratinocyte carcinoma (KC) but predictive factors and their risk rates are unknown.

Objectives: To develop and internally validate a prediction model to calculate the absolute risk of a first KC in AK patients.

Methods : The risk prediction model was based on the prospective population-based Rotterdam Study cohort. We hereto analyzed data of participants with at least 1 AK-lesion at cohort baseline using a multivariable Cox proportional hazards model and included 13 a priori defined candidate predictor variables considering phenotypic, genetic and lifestyle risk factors. KCs were identified by linkage of the data with the Dutch Pathology Registry.

Results: Of the 1,169 AK-participants at baseline, 176 (15.1%) developed a KC after a median follow-up of 1.8 years. The final model with significant predictors was obtained after backward stepwise selection and comprised the presence of 4-9 AKs (hazard ratio (HR): 1.68, 95% confidence interval (CI): 1.16-2.42), 10 or more AKs (HR: 2.43, 95% CI: 1.64-3.61), AK-localization on upper extremities (HR: 0.75, 95% CI: 0.52-1.08) or elsewhere except the head (HR: 1.40, 95% CI: 0.98-2.01), and coffee consumption (HR: 0.92, 95% CI: 0.84-1.01). Evaluation of the discriminative ability of the model showed a bootstrap validated c-index of 0.60.

Conclusions: This is the first multivariable risk prediction model for KC development in patients with AKs. We showed that the risk of KC can be calculated with four easily assessable predictor variables. Given the c-index, extension of the model with additional, currently unknown predictor variables is desirable.

INTRODUCTION

Actinic keratoses (AKs) are premalignant lesions and can be considered a clinical biomarker for cutaneous photodamage (1). Population-based studies report a high prevalence of AKs, especially in elderly people of European ancestry (2, 3). In the Netherlands, 23.5% of the population aged 50 years or older has one or multiple AKs (4). Individual AKs may progress into cutaneous squamous cell carcinoma (cSCC). Additionally, as a marker of ultraviolet radiation (UVR) induced DNA damage, the presence of AK is a risk factor for keratinocyte carcinoma (KC) in general, including basal cell carcinoma (BCC) (5-7). It is however unclear which AK patients will develop KCs and how high this risk rate is, although several AK-characteristics such as the presence of multiple AKs and their anatomical site as well as general phenotypic factors (e.g. light pigment status) and exposure related items (e.g. high UVR-exposure) have been described to increase progression risk (8-10). Correctly identifying high risk patients is important to detect KCs in an early stage and to ensure timely intervention. Moreover, stratified AK-management may reduce patients' anxiety, provide better management for high risk individuals, and optimize the use of healthcare resources (11).

Until now, several KC prediction models have been developed regarding the occurrence of either a first or subsequent KC in the general population (12-15). However, none of these assessed what factors predict a KC in an AK-population, which is a very relevant question for many healthcare providers. We therefore aimed to develop a simple model to predict the absolute risk of a first KC in patients with AK, taking into account phenotypic, lifestyle and genetic susceptibility factors, by analyzing over 1,000 AK-participants from the prospective population-based Rotterdam Study cohort (RS).

PATIENTS AND METHODS

Study population

The RS is a prospective population-based cohort study comprising 14,926 participants of 45 years and older from the general population of Ommoord in Rotterdam, the Netherlands. From July 1989 to present, the participants undergo regular examinations in a research facility and interviews are conducted at home about every 3-4 years. Between 2010 and 2016, complete skin examinations were performed during the RS routine, focusing on common skin diseases including AK as well as potential risk factors. We included participants with at least 1 AK-lesion during one of these examinations in our model. The date of first AK diagnosis in the RS cohort served as the starting point of follow-up.

The RS has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population

Screening Act WBO, license number 1071272-159521-PG). All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians. Details of the study design and objectives have been described before (16).

Case definition

The study outcome was defined as a first KC, either BCC or cSCC, after AK diagnosis. To identify all KC cases, the RS participants were linked to the Dutch nationwide network and registry of histo- and cytopathology in the Netherlands (PALGA) using encrypted patient data (combination of the patient's gender, birth date and first four to eight letters of the (maiden) family name). Participants with a KC diagnosis prior to their AK diagnosis were excluded, as our study was focused on skin cancer naïve AK patients. Follow-up of all participants ended at the time of KC diagnosis or, when this outcome measure was not met, at the date of censoring. Censoring events were death as assessed from the municipal register or end of available PALGA follow-up on July 31st 2018, whichever occurred first.

Candidate predictor variables

The candidate predictor variables were a priori selected based on literature review and clinical expertise and were categorized as follows: AK specific variables, phenotypic factors, lifestyle factors, and a genetic susceptibility variable.

As AK specific variables, we included the number of AKs at diagnosis (1-3, 4-9, >10) and categorized the location of AKs into 3 main groups: head, upper extremities and elsewhere. In case of AKs on multiple locations per participant, more than one location variable could be selected.

We included four phenotypic factors, namely age at AK diagnosis in the RS (years), sex, tendency to develop sunburn and pigment status. The latter constituted of a combination of hair and eye color when young, as reported previously (13).

As lifestyle factors, smoking and coffee consumption were included in cups per day. Regarding UVR-exposure, we selected variables reflecting intermittent or chronic exposure to UVR. Intermittent UVR exposure was defined as a combination of likeliness to be outdoors when the sun is shining/having mainly outside hobbies, going on holidays to a sunny country at least 4 weeks per year and sun bed usage for at least 10 times in the past 5 years. Chronic UVR-exposure was assessed as the history of occupational outdoor work for at least 4 hours per day during at least 25 years.

We calculated a genetic risk score (GRS) per AK patient by retrieving 7 SNPs that were significantly associated with both BCC and cSCC occurrence from the most recent genome-

wide association studies (GWAS) (17, 18) (Table S1 in Supporting Information). A detailed description of the GRS computation method is presented in the Supporting Information online. In brief, a weighted GRS was calculated using the regression coefficients of published associations between the selected SNP and cSCC (18). The genetic scores were computed as follows: GRS= $\sum \beta iGi$; where βi is the log(Odds Ratio) of the SNP and *G*i is the number of per-SNP risk alleles (0, 1 or 2).

All predictor variables were measured at baseline i.e. at the moment of AK-diagnosis, DNA from whole blood was extracted at the start of each cohort (I-III) within RS. For lifestyle and UVR-exposure variables, values from an earlier examination round were used if they were missing at baseline.

Model development and performance

We used a Cox proportional hazards model to determine the probability of first KC development in patients with AK, taking censoring into account. Before starting the model development, collinearity among plausible categorical predictor variables was tested with Cramer's V statistic with no evidence found for multicollinearity. We imputed all missing values except for GRS 10 times using multivariate imputation by chained equations (MICE) (19), under the assumption that the data were missing at random. We included all candidate predictors, the outcome (KC or censored) and the follow-up time in years in the imputation model. Also, RS cohort number (I-III) and socioeconomic status of the participants were included as auxiliary variables.

Univariable analyses were performed for all candidate predictor variables and the occurrence of KC. For the continuous variables age and coffee consumption, we explored a possible non-linear relationship using a natural cubic spline with two degrees of freedom. The use of a spline for these variables did neither significantly improve the fit of our model (measured with the X2-value) nor provide graphical evidence for a non-linear relationship. We therefore included these variables in their linear forms.

Regardless of their p-values in the univariable analyses, all candidate predictors were included in the multivariable model (20, 21). We reduced the multivariable model by backward stepwise selection using a liberal p-value of 0.20 for the inclusion of predictors to reduce selection bias and optimism (21). The estimated regression coefficients and variances from the 10 imputed datasets were combined based on Rubin's rules (22).

We assessed the predictive performance of our model in terms of discrimination using Harrell's concordance index (c-index). The c-index in survival context can be interpreted as the probability that the model assigns a higher predicted risk of KC development to a patient (from a randomly chosen pair of patients) that develops KC earlier in time compared

to a patient developing KC later in time and varies from 0.5 (non-informative model) to 1.0 (perfect model) (23). As a means of internal validation, we used bootstrapping to correct the c-index for optimism.

To account for overfitting, we multiplied the regression coefficients from our final model with a shrinkage factor, which we estimated with bootstrapping (1000 replications). Shrinkage of regression coefficients towards average is meant to improve predictions in future patients by preventing extreme distributions of the predictions (21).

A sensitivity analysis was performed excluding all participants with missing values. Reporting of the model is done according to the TRIPOD Statement (24).

Model presentation

To provide individualized predictions on the risk of first KC development in AK patients, we made a risk prediction tool based on the shrunk regression coefficients of our internally validated model using Microsoft Excel (2010).

Descriptive statistics were computed using IBM SPSS Statistics for Windows, version 24.0. (Armonk, NY: IBM Corp.). Model development and internal validation were conducted using R statistical software version 3.5.0 (R Foundation for Statistical Computing, Vienna, Austria) with the mice, Hmisc and rms libraries.

RESULTS

Study population

Selection of all participants with at least 1 AK-lesion at baseline resulted in 1,558 subjects. After linkage with PALGA, 389 participants were excluded who had at least 1 KC prior to their AK diagnosis. The median follow-up of the remaining 1,169 participants was 5.2 years (interquartile range (IQR) 3.5-6.9), during which 176 participants developed a KC at a median follow-up of 1.8 years (IQR 0.2-3.8). The majority of participants (58.9%) had 1-3 AK-lesions at baseline, mainly located on the head (84.4%). The overall median age was 73.0 years (IQR 67.0-80.0) and 55% of all participants were men (Table 1).

Table 1. Descriptive characteristics of the 1,169 participants with at least 1 actinic keratosis (AK) at baseline and keratinocyte carcinoma (KC) cases (N=176) separately

Candidate predictor variables	Category	Overall (N=1,169)	KC cases (N=176)
Number of participants		1,169 (100%)	176 (15.1%)
Follow-up time (years)	Median (IQR)	5.2 (3.5-6.9)	1.8 (0.2-3.8)
Age at AK-diagnosis (years)	Median (IQR)	73.0 (67.0-80.0)	73.0 (67.0-79.0)
Sex	Male	643 (55.0%)	96 (54.5%)
Number of AKs at diagnosis	1-3	689 (58.9%)	78 (44.3%)
	4-9	290 (24.8%)	49 (27.8%)
	≥10	190 (16.3%)	49 (27.8%)
AK on the head[1]	No	182 (15.6%)	26 (14.8%)
	Yes	987 (84.4%)	150 (85.2%)
AK on upper extremities[2]	No	882 (75.4%)	132 (75.0%)
	Yes	287 (24.6%)	44 (25.0%)
AK on other locations[3]	No	973 (83.2%)	132 (75.0%)
	Yes	196 (16.8%)	44 (25.0%)
Pigment status[4]	Dark	222 (19.0%)	32 (18.2%)
	Intermediate	618 (52.9%)	95 (54.0%)
	Light	281 (24.0%)	43 (24.4%)
	Missing	48 (4.1%)	6 (3.4%)
Being easily sunburned	No	704 (60.2%)	100 (56.8%)
	Yes	416 (35.6%)	69 (39.2%)
	Missing	49 (4.2%)	169 (96.0%)
Intermittent sun exposure[5]	No	114 (9.8%)	18 (10.2%)
	Yes	732 (62.6%)	97 (55.1%)
	Missing	323 (27.6%)	61 (34.7%)
Outdoor work[6]	No	462 (39.5%)	74 (42.0%)
	Yes	133 (11.4%)	20 (11.4%)
	Missing	574 (49.1%)	82 (46.6%)
Smoking	Never	357 (30.5%)	50 (28.4%)
	Current or ever	798 (68.3%)	123 (69.9%)
	Missing	14 (1.2%)	3 (1.7%)
Coffee consumption (cups/day)	Median (IQR)	3.3 (1.4-3.3)	1.4 (1.4-3.3)
	Missing	131 (11.2%)	23 (13.1%)
GRS	Median (IQR)	1.0 (1.0-1.1)	1.1 (1.0-1.1)
	Missing	159 (13.6%)	25 (14.2%)

AK, actinic keratosis; KC, keratinocyte carcinoma; UVR, ultraviolet radiation; IQR, interquartile range; GRS, genetic risk score.

[1]Presence of AK on the face, ears and/or scalp. [2]Presence of AK on the back of the hands and/or forearms.
[3]Presence of AK on locations elsewhere (not specified). [4]A combination of hair- and eye color when young.

[5]Combination variable of a confirmatory answer to one or more of the following questions:
Are you likely to be outside when the sun is shining/do you mainly have outside hobbies?
Do you go on holidays to a sunny country at least 4 weeks per year on average?
Have you used a sunbed for at least 10 times during the past 5 years?

[6]To have been/worked outdoors for at least 4 hours daily during at least 25 years.

Predictors for a first KC

In univariable analyses, the presence of 4-9 AKs and 10 or more AKs, an AK-localization outside the head or upper extremities and increasing age were significantly associated with a higher risk of KC development (Table 2). On the contrary, the risk of KC occurrence decreased per cup of coffee consumption (hazard ratio (HR): 0.92, 95% confidence interval (CI): 0.84-1.01). After backward stepwise selection, four predictor variables remained in the final model: number of AKs at diagnosis (either 4-9 or 10 or more), localization of AKs on the upper extremities, localization of AKs elsewhere except on the head, and coffee consumption. After adjustment for all other predictors in multivariable analysis, age was not significantly associated with KC anymore. Having 10 or more AKs was the strongest predictor with a 2.5 times higher hazard of KC development compared to the presence of 1-3 AKs (HR: 2.47, 95% CI: 1.65-3.61). Although evidence exists for a familial aggregation basis of skin cancer (17, 18, 25-27), the GRS based on SNPs associated with KC did not increase the risk of KC development in our AK population.

A sensitivity analysis on 335 participants with no missing values yielded comparable HRs and the same reduced multivariable model (data not shown).

The overall apparent c-index of the final model was 0.61 (95% CI: 0.56-0.66). After internal validation of the model with bootstrapping, the optimism corrected c-index reduced to 0.60

Model presentation

Figure 1 shows an image of the risk prediction tool that can be used easily to predict an AK patient's risk of first KC development, given the four prognostic factors from the final model. The regression coefficients of these predictors have been multiplied with an estimated shrinkage factor of 0.91. After filling in the individual values for each of these predictors, the tool calculates the percentage risk of a first KC in 1, 3 and 5 years. For example, a patient with 10 AKs spread over the upper extremity and other body sites except the head and who drinks 3 cups of coffee per day, has a 22% risk of KC development in 5 years. The Excel file containing this risk prediction tool is available for reference in the Supporting Information.

Table 2. Associations (hazard ratios (HRs) with confidence intervals (CIs)) between candidate predictor variables and development of a first KC (n=176) using a Cox proportional hazards model

Candidate predictor variables	Coding	Univariable HR (95% CI)	Multivariable HR[1] (95% CI)
Age		1.01 (0.99-1.03)*	-
Sex	Female	1.03 (0.77-1.39)	-
Number of AKs at diagnosis	1-3	Reference	Reference
	4-9	1.59 (1.11-2.28)**	1.68 (1.17-2.42)**
	≥10	2.47 (1.73-3.53)***	2.44 (1.65-3.61)***
AK on the head[2]	Yes	1.09 (0.72-1.65)	-
AK on upper extremities[3]	Yes	0.99 (0.71-1.41)	0.75 (0.52-1.08)*
AK on other locations[4]	Yes	1.72 (1.23-2.43)***	1.40 (0.98-2.01)*
Pigment status[5]	Dark	Reference	-
	Intermediate	1.01 (0.68-1.51)	
	Light	1.00 (0.63-1.57)	
Being easily sunburned	Yes	1.11 (0.82-1.51)	-
Intermittent sun exposure[6]	Yes	0.84 (0.52-1.36)	-
Outdoor work[7]	Yes	0.93 (0.58-1.51)	-
Smoking	Ever	1.09 (0.78-1.51)	-
Coffee consumption (cups/day)		0.92 (0.84-1.01)*	0.92 (0.84-1.01)*
GRS		1.92 (0.58-6.31)	-

HR, hazard ratio; CI, confidence interval; AK, actinic keratosis; UVR, ultraviolet radiation; GRS, genetic risk score.

*P-value <0.20, ** P-value <0.05, and *** P-value <0.005.

[1]Final model after backward stepwise selection. [2]Presence of AK on the face, ears and/or scalp. [3]Presence of AK on the back of the hands and/or forearms. [4]Presence of AK on locations elsewhere (not specified).
[5]A combination of hair- and eye color when young. [6]Combination variable of a confirmatory answer to one or more of the following questions:
Are you likely to be outside when the sun is shining/do you mainly have outside hobbies?
Do you go on holidays to a sunny country at least 4 weeks per year on average?
Have you used a sunbed for at least 10 times during the past 5 years?
[7]To have been/worked outdoors for at least 4 hours daily during at least 25 years.

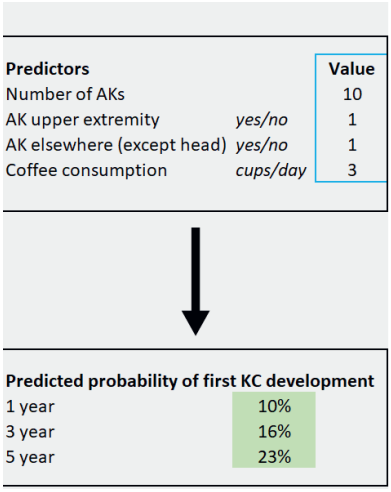

Fig 1. Risk-prediction tool for KC development in patients with AK, filled in for an example patient with 10 AKs, located on the upper extremity and elsewhere (not on the head), and who drinks three cup sof coffee per day. The subsequent formula is used to predict the percentage risk of a first KC at 1 year after AK diagnosis: P = [1–(EXP(–EXP(lp–lp.centered)*baselinehaz))] 9 100% where lp = –0.278*AK location upper extremity + 0.345*AK location elsewhere except head –0.060*cups of coffee per day + presence of multiple AKs (0 if 1–3 AKs,0.515 if 4–9 AKs, 0.888 if ≥ 10 AKs), lp.centered = 0.104 and the baseline hazard is 0.057. Both lp and lp.centered have been multiplied by the shrinkage factor of 0.91. For the risks at 3 and 5 years, the baseline hazard should be replaced by 0.092 and 0.144, respectively

DISCUSSION

Our population-based study with over 1,000 participants provides the first risk prediction model for an AK specific patient group and encompasses readily available phenotypic, lifestyle and genetic KC susceptibility factors. The strongest predictor of a first KC was having 10 or more AKs at diagnosis, which increased the KC risk with 2.5 fold. This is in line with other cohort studies demonstrating a strong dose-response relationship between the number of AKs and the risk of a KC (9, 28-30). This finding could be explained through several theories. Firstly, cumulative UVR-exposure underlies both AK and KC development. A study of the association between AKs and KCs showed that the etiologic factors for AK development were essentially equal to the etiologic factors for both BCC and cSCC development (31). Secondly, AKs can be seen as an early phase in the biologic continuum that eventually culminates in cSCC, which means that part of the AKs in our cohort might have progressed directly to cSCC (30). Thirdly, from the concept of field cancerization, the presence of multiple AKs forms the ultimate groundwork for the progression of epithelial carcinogenesis (32).

Little is known about the risk of KC development based on AK affected body site. We found that AKs localized on the upper extremities significantly decreased and AKs localized outside the head and upper extremity regions significantly increased the risk of KC. This finding is

consistent with a Dutch systematic review concluding that patients with AKs on the head or upper extremities are less likely to develop KCs compared to patients with AKs on the neck, trunk or lower extremities (33). An explanation for our finding is not straightforward. It is remarkable that covered body sites showed higher risk rates than the more chronic sun exposed head and upper extremity regions, which may hint to a different carcinogenesis pattern than chronic UVR-exposure.

Coffee consumption is a much discussed factor in the field of skin cancer carcinogenesis. In our analyses, we found that coffee consumption significantly reduced the risk of a first KC with 8% per cup of coffee. Findings from mainly laboratory and animal studies have indicated a possible protective effect of caffeine against KC development through induction of apoptosis in UVR damaged keratinocytes (34, 35) as well as inhibition of UVR induced carcinogenesis (36). The chemo-protective effect of caffeine for KC (especially for BCC) in European-descent populations has recently been supported by two meta-analyses of observational studies as well (37, 38). Additionally, coffee intake can be considered a proxy for good health and wellbeing as consumers of coffee often have a healthier lifestyle in general (39).

Remarkably, none of the UVR related predictor variables nor participants' pigment status were associated with KC. This is in line with other KC prediction models that used the same or comparable sun exposure variables (7, 13-15). Since we selected our study population on the presence of AKs, which in a way can be considered primary KCs because of equal risk profiles, index-event bias may underlie the results (40): UVR-exposure is a pivotal risk factor for the occurrence of AKs, but in our model paradoxically not for a subsequent KC. This is because conditioning on the presence of AKs generates dependence between all other known and unknown risk factors, eventually leading to underestimated or even reversed effects and biasing the risk rates towards the null. We indeed found HRs that were low (for being easily sunburned) or even seemed to be protective (history of outdoor work and intermittent sun exposure) in our univariable analyses, which are likely to be caused by index event bias.

Limitations

With the current internally validated discriminative value, our risk stratification tool might not be clinically useful yet. Although we were able to include all variables of interest as derived from literature and clinical expertise, we found a c-index of 0.60. This poor to moderate c-index could be explained by the very homogeneous nature of our study population. AK patients are a priori people with a fair skin, at age, and who have all had cumulative UVR-exposure throughout the years. Finding additional KC predictors that specifically discriminate within the AK population is therefore a challenging task and the phenotypic, lifestyle and genetic risk factors at hand appeared to be insufficient. Another

explanation for the moderate c-index might be that we have not separated BCC and cSCC as separate outcome measures due to insufficient power (17 degrees of freedom in total). Effect estimates per predictor could differ for BCC and cSCC, hereby influencing the discriminative ability of our model. However, a quick subgroup check on univariable analyses between the predictors and BCC/cSCC separately did not show any differences between both KC types (data not shown). Still, given the very limited existing knowledge in the AK prognostic field, we believe that the current model provides important insights and can be used to build upon for more extensive models and the selection of tailored variables. Another limitation is that we only assessed the number of AKs at moment of diagnosis during RS, while this could have fluctuated during follow-up due to e.g. treatment or spontaneous regression of the lesion. However, as we assessed the overall risk of KC development considering all AKs in a patient instead of the lesion specific progression risk, we do not expect that potential slight changes in the number of AKs would have affected the risk rates nor the c-index of our model.

Conclusion

Risk of first KC development in AK patients can be predicted by a simple tool including the number and location of AKs along with coffee consumption. This information can help physicians in identifying high risk AK patients and planning further AK management. Extension with additional predictive factors and external validation thereafter are needed before use in clinical practice is recommended.

References

1. Saladi RN, Persaud AN. The causes of skin cancer: a comprehensive review. Drugs Today (Barc). 2005;41(1):37-53.
2. Naldi L, Chatenoud L, Piccitto R, Colombo P, Placchesi EB, La Vecchia C, et al. Prevalence of actinic keratoses and associated factors in a representative sample of the Italian adult population: Result from the Prevalence of Actinic Keratoses Italian Study, 2003-2004. Arch Dermatol. 2006;142(6):722-6.
3. Harvey I, Frankel S, Marks R, Shalom D, Nolan-Farrell M. Non-melanoma skin cancer and solar keratoses II analytical results of the South Wales Skin Cancer Study. Br J Cancer. 1996;74(8):1308-12.
4. Flohil SC, van der Leest RJ, Dowlatshahi EA, Hofman A, de Vries E, Nijsten T. Prevalence of actinic keratosis and its risk factors in the general population: the Rotterdam Study. J Invest Dermatol. 2013;133(8):1971-8.
5. Vatve M, Ortonne JP, Birch-Machin MA, Gupta G. Management of field change in actinic keratosis. The British journal of dermatology. 2007;157 Suppl 2:21-4.
6. Curtius K, Wright NA, Graham TA. An evolutionary perspective on field cancerization. Nat Rev Cancer. 2018;18(1):19-32.
7. Foote JA, Harris RB, Giuliano AR, Roe DJ, Moon TE, Cartmel B, et al. Predictors for cutaneous basal- and squamous-cell carcinoma among actinically damaged adults. Int J Cancer. 2001;95(1):7-11.
8. Belbasis L, Stefanaki I, Stratigos AJ, Evangelou E. Non-genetic risk factors for cutaneous melanoma and keratinocyte skin cancers: An umbrella review of meta-analyses. J Dermatol Sci. 2016;84(3):330-9.
9. Khalesi M, Whiteman DC, Doi SA, Clark J, Kimlin MG, Neale RE. Cutaneous markers of photo-damage and risk of Basal cell carcinoma of the skin: a meta-analysis. Cancer Epidemiol Biomarkers Prev. 2013;22(9):1483-9.
10. Richard MA, Amici JM, Basset-Seguin N, Claudel JP, Cribier B, Dreno B. Management of actinic keratosis at specific body sites in patients at high risk of carcinoma lesions: expert consensus from the AKTeam of expert clinicians. J Eur Acad Dermatol Venereol. 2018;32(3):339-46.
11. Hollestein LM, de Vries E, Aarts MJ, Schroten C, Nijsten TE. Burden of disease caused by keratinocyte cancer has increased in The Netherlands since 1989. Journal of the American Academy of Dermatology. 2014;71(5):896-903.
12. Smedinga H, Verkouteren JAC, Steyerberg EW, Hofman A, Nijsten T, Vergouwe Y. Occurrence of metachronous basal cell carcinomas: a prognostic model. The British journal of dermatology. 2017;177(4):1113-21.
13. Verkouteren JAC, Smedinga H, Steyerberg EW, Hofman A, Nijsten T. Predicting the Risk of a Second Basal Cell Carcinoma. J Invest Dermatol. 2015;135(11):2649-56.
14. Whiteman DC, Thompson BS, Thrift AP, Hughes MC, Muranushi C, Neale RE, et al. A Model to Predict the Risk of Keratinocyte Carcinomas. J Invest Dermatol. 2016;136(6):1247-54.
15. Wang W, Jorgenson E, Ioannidis NM, Asgari MM, Whittemore AS. A Prediction Tool to Facilitate Risk-Stratified Screening for Squamous Cell Skin Cancer. J Invest Dermatol. 2018;138(12):2589-94.

16. Ikram MA, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2018 update on objectives, design and main results. European journal of epidemiology. 2017;32(9):807-50.
17. Chahal HS, Lin Y, Ransohoff KJ, Hinds DA, Wu W, Dai HJ, et al. Genome-wide association study identifies novel susceptibility loci for cutaneous squamous cell carcinoma. Nat Commun. 2016;7:12048.
18. Sordillo JE, Kraft P, Wu AC, Asgari MM. Quantifying the Polygenic Contribution to Cutaneous Squamous Cell Carcinoma Risk. J Invest Dermatol. 2018;138(7):1507-10.
19. Van Buuren S. Flexible Imputation of Missing Data. Boca Raton, FL, USA: Chapman and Hall/CRC; 2012. pp 342 p.
20. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. BMJ. 2009;338:b604.
21. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating.: Springer Science+Business Media, LLC: NY, USA; 2009.
22. Vergouwe Y, Royston P, Moons KG, Altman DG. Development and validation of a prediction model with missing predictor data: a practical approach. Journal of clinical epidemiology. 2010;63(2):205-14.
23. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-87.
24. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
25. Hussain SK, Sundquist J, Hemminki K. The effect of having an affected parent or sibling on invasive and in situ skin cancer risk in Sweden. J Invest Dermatol. 2009;129(9):2142-7.
26. Hemminki K, Zhang H, Czene K. Familial invasive and in situ squamous cell carcinoma of the skin. Br J Cancer. 2003;88(9):1375-80.
27. Chahal HS, Wu W, Ransohoff KJ, Yang L, Hedlin H, Desai M, et al. Genome-wide association study identifies 14 novel risk alleles associated with basal cell carcinoma. Nat Commun. 2016;7:12510.
28. Kricker A, Armstrong BK, English DR, Heenan PJ. Pigmentary and cutaneous risk factors for non-melanocytic skin cancer--a case-control study. Int J Cancer. 1991;48(5):650-62.
29. Armstrong BK, Kricker A. The epidemiology of UV induced skin cancer. J Photochem Photobiol B. 2001;63(1-3):8-18.
30. Salasche SJ. Epidemiology of actinic keratoses and squamous cell carcinoma. Journal of the American Academy of Dermatology. 2000;42(1 Pt 2):4-7.
31. Marks R, Rennie G, Selwood T. The relationship of basal cell carcinomas and squamous cell carcinomas to solar keratoses. Arch Dermatol. 1988;124(7):1039-42.
32. Torezan LA, Festa-Neto C. Cutaneous field cancerization: clinical, histopathological and therapeutic aspects. An Bras Dermatol. 2013;88(5):775-86.

33. Smit P, Plomp E, Neumann HA, Thio HB. The influence of the location of the lesion on the absolute risk of the development of skin cancer in a patient with actinic keratosis. J Eur Acad Dermatol Venereol. 2013;27(6):667-71.
34. Heffernan TP, Kawasumi M, Blasina A, Anderes K, Conney AH, Nghiem P. ATR-Chk1 pathway inhibition promotes apoptosis after UV treatment in primary human keratinocytes: potential basis for the UV protective effects of caffeine. J Invest Dermatol. 2009;129(7):1805-15.
35. Han W, Ming M, He YY. Caffeine promotes ultraviolet B-induced apoptosis in human keratinocytes without complete DNA repair. J Biol Chem. 2011;286(26):22825-32.
36. Huang MT, Xie JG, Wang ZY, Ho CT, Lou YR, Wang CX, et al. Effects of tea, decaffeinated tea, and caffeine on UVB light-induced complete carcinogenesis in SKH-1 mice: demonstration of caffeine as a biologically important constituent of tea. Cancer Res. 1997;57(13):2623-9.
37. Caini S, Cattaruzza MS, Bendinelli B, Tosti G, Masala G, Gnagnarella P, et al. Coffee, tea and caffeine intake and the risk of non-melanoma skin cancer: a review of the literature and meta-analysis. European journal of nutrition. 2017;56(1):1-12.
38. Vaseghi G, Haghjoo-Javanmard S, Naderi J, Eshraghi A, Mahdavi M, Mansourian M. Coffee consumption and risk of nonmelanoma skin cancer: a dose-response meta-analysis. Eur J Cancer Prev. 2018;27(2):164-70.
39. Mirza SS, Tiemeier H, de Bruijn RF, Hofman A, Franco OH, Kiefte-de Jong J, et al. Coffee consumption and incident dementia. European journal of epidemiology. 2014;29(10):735-41.
40. Dahabreh IJ, Kent DM. Index event bias as an explanation for the paradoxes of recurrence risk research. JAMA. 2011;305(8):822-3.
41. Fortier I, Raina P, Van den Heuvel ER, Griffith LE, Craig C, Saliba M, et al. Maelstrom Research guidelines for rigorous retrospective data harmonization. Int J Epidemiol. 2017;46(1):103-5.

Part IV

General Discussion

Chapter 8

General Discussion

GENERAL DISCUSSION

The overall aim of this thesis was to evaluate ethics, cost-effectiveness, and prediction modeling of cancer screening and treatment. We addressed four research questions. The main findings on these research questions are shown in box 1. In this chapter, the results of our studies will be discussed in the light of application in screening and the role of prediction models, followed by future directions and recommendations.

Box 1 - Overview of main findings per research question

1. How can we promote autonomous choices in informed consent procedures regarding participation in epigenetic risk-tailored cancer screening?
Informed consent models that inform patients in different stages with varying levels of information (e.g. the tiered-layered staged informed consent model) are highly suitable for epigenetic risk-stratified screening programs. We developed a framework to guide the development of Patient Decision Aids (PDAs), to support informed consent and promote autonomous choices in the specific context of epigenetic cancer screening programs.

2. What is the cost-effectiveness of offering prophylactic hysterectomy to first-degree relatives of patients with colorectal cancer and Lynch Syndrome?
Offering prophylactic hysterectomy to 40-80 year old women with Lynch Syndrome is expected to add 0.5 QALY per person at acceptable costs. Women may decide to have prophylactic hysterectomy at age 35 years, depending on their individual disutility for prophylactic hysterectomy and premature menopause.

3. What is the quality of risk prediction models for endometrial cancer in the general population?
Only a few models have been developed to predict endometrial cancer in asymptomatic or symptomatic women. The usefulness of most models is unclear considering methodological shortcomings, such as the handling of missing data and the selection of predictors.

4. How can we obtain accurate and valid individual estimates of progression from premalignant to malignant cancer based on prediction models to guide decision making?
Predicting progression of precancer lesions to clinical cancer is possible based on relevant clinical parameters. Prediction models might be able to accurately distinguish between patients with a high and low risk of progression. However, external validation is currently lacking, which is a vital step before implementing these models in clinical practice.

Cancer screening

Patient autonomy in cancer screening

Technological advances have made information on individuals' (epi)genome relatively easily accessible compared to ten years ago. Such genetic information can be used to enrich cancer screening programs, as the (epi)genome holds valuable information on the presence of DNA methylations due to external factors (i.e. smoking status, (former) use of contraceptives) or genetic germline mutations (i.e. Lynch Syndrome or BRCA 1/2). The patient autonomy might be endangered if results from the (epi)genome are incorporated in cancer screening programs, as it is questionable if patients are capable of making an informed choice. In line with previous literature, we found that patient decision aids (PDAs) might support patients in making an autonomous choice regarding participation in a cancer screening program (Chapter 2).[1] Over the past decade, the number of developed PDAs has increased substantially.[2] Although the design and graphical presentation of each PDA may vary, every PDA should at least include a section with information on the test and risk probabilities, followed by a value clarification method.[3] The latter will help patients to identify their values regarding the decisions they are facing. For instance, a value clarification method in a PDA for breast cancer screening helps women to identify their feelings regarding risk and benefits of mammography screening.[4]

Figure 1 – Example of a value clarification method in a Patient Decision Aid (PDA) for breast cancer screening[4]

Several studies have investigated the clinical role of PDAs and have shown that they are effective in terms of improving knowledge, increasing patient participation and reducing decisional conflict.[2, 5] However, the effectiveness of PDAs remains limited if they are

developed incorrectly. As a consequence, these PDAs may cause harm instead of supporting patient autonomy.

The development process of PDAs involves different steps, ranging from setting the scope and aim, to presenting the final PDA. While the development process of most PDAs includes the first and last step of development, the intermediary steps are often overlooked.[6] These intermediary steps include field testing a test version of the PDA with patients and physicians and testing the final version in a real-life setting. The inclusion of patients and physicians in the development process is of great importance, as it increases the chance that the PDA suits to the preferences of the targeted patient population. Furthermore, barriers to using the PDA for physicians in clinical practice have been identified early, which eases the implementation process.

While correctly developed PDAs may aid in the process of deriving an autonomous choice on participating in a cancer screening program, other barriers for the implementation of epigenetic risk-tailored cancer screening continue to exist. These barriers are multidimensional and contain both practical and ethical issues (Box 2).[7] Some of these barriers might be rather easy to overcome, but some might have more consequence and should therefore be discussed further. For example, the ethical barrier that the results might contain sensitive information on (past) behavior may provoke a societal debate on the handling and storage of this type of information. Improper use by third parties may have negative consequences, for example on access to health insurers, and could also lead to societal pressure because we can hold a person responsible for his high risk because of past behavior.

The process of implementing risk-tailored cancer screening program will therefore be challenging. The question remains if such innovations will be implemented in nationwide screening programs in the near future. Technological advances such as analysis of the epigenome could enrich existing cancer screening programs, if implemented correctly. This might result in more individualized risk predictions, which suits the movement towards personalized medicine and better cost-effectiveness of screening by better targeting screening at those at increased risk.

Box 2 – Barriers and potential solutions for implementation of epigenetic risk-tailored cancer screening

Practical issues

- Current invitations are mostly sent by mail, without the possibility to contact a physician.
  Solution: invitations should be revised and updated with new information on the changes in the cancer screening program and the possibility to contact a physician should be adopted.

- Current policies on receiving the results might be insufficient in the new situation, since the results will contain more information that might provoke questions
  Solution: offer the possibility to obtain the results via a physician, for example a consult with the general practitioner.

- Results of the screening program might also have consequence for family members – how will they be informed?
  Solution: optional visits to a genetic counselor should be made possible for participants of the screening program.

Ethical issues

- Current informed consent procedures endanger patient autonomy.
  Solution: informed consent procedures should be revised and supported by patient decision aids to ensure that participants make an informed choice.

- Results from the screening program might contain sensitive information, as it contains information on past (harmful) behavior.
  Solution: it should be determined who has access to these data to prevent improper use by third parties (e.g. health insurers).

Cost-effectiveness of prophylactic surgery

Cost-effectiveness studies are important to inform patients, physicians and policymakers on the costs and effects of interventions, including screening strategies. Screening strategies are not solely targeted at seemingly healthy individuals, but could also be designed for individuals with an increased risk of developing cancer. For example, high-risk individuals might benefit from prophylactic surgery such as prophylactic mastectomy in BRCA 1/2 mutation carriers or prophylactic hysterectomy in women with Lynch Syndrome. In Chapter 3 we studied the cost-effectiveness of offering prophylactic hysterectomy in first-degree female relatives with Lynch Syndrome of probands with colorectal cancer. We found that prophylactic hysterectomy is most cost-effective when offered to women aged 40-80

years. However, the decision to undergo prophylactic surgery remains a preference-sensitive decision, and depends on individuals' perception and disutility of prophylactic surgery and its consequence such as premature menopause. These disutilities are likely to affect the gains (e.g. QALY gained) of prophylactic surgery. Several factors influence the utility of participating in cancer screening programs, either with or without prophylactic treatment (Table 1).

Table 1. Factors that influence the probability of participating in cancer screening programs

Factor	Effect
Negative feelings on screening (anxiety, stress)	Negative
Risk of developing cancer that is screened for in the screening program	Low risk: negative High risk: positive
Associated negative side-effects of participation in screening (e.g. consequences of prophylactic surgery)	Negative
Risk reduction due to prophylactic surgery	Positive
Life expectancy of participant	Low life expectancy: negative High life expectancy: positive
Rate of overdiagnosis	Low rate: positive High rate: negative
False-positive rate	Low rate: positive High rate: negative

A recent study confirmed that patients' preferences influence the level of overall gain from participating in cancer screening.[8] Based on a microsimulation study, the lifetime quality adjusted life-years (QALYs) gained from low-dose computed tomography (LDCT) screening for lung cancer varied substantially when different patient preferences where compared. For example, a significant shift from net benefit to net harm was found when unfavorable preferences for screening were compared with favorable preference for screening, while maintaining the same lung cancer risk in both groups. The difference in preferences also resulted in a substantial increase in the number needed to screen to avoid one lung cancer death.[8] Although these results are simulated for people who attend lung cancer screening with LDCT, without the consequence of prophylactic surgery, these results are most likely also relevant in the case of offering prophylactic surgery to Lynch positive women. Hence, it must be acknowledged that individual (dis)utilities probably influence women in their decision to undergo prophylactic surgery and that a formal cost-effectiveness analysis might not fully reflect real life decisions and the delicate balance between benefits and harms.

As individual preferences might influence the decision to undergo prophylactic surgery, it is no surprise that the uptake of prophylactic surgery in high risk populations varies

between individuals and countries.[9-11] For example, the uptake of contralateral prophylactic mastectomy in high risk women varied from 0% in Norway to 49% in the United States.[9] However, if one decides to undergo prophylactic surgery, the overall satisfaction with the decision is high. The satisfaction rates for prophylactic mastectomy and oophorectomy vary from 87-88% and 89-95% respectively. [12, 13] Furthermore, prophylactic surgery reduces cancer specific anxiety and psychological morbidity in patients who decide to undergo prophylactic surgery compared to patients who decline it.[14] These results leave us with the question whether the low initial uptake of prophylactic surgery truly reflects individuals' preference for not participating, or that there might be other factors that lower the chance of having prophylactic surgery. Age, prior cancer and previous utilization of prophylactic surgery were significant predictors of uptake of prophylactic surgery in high risk women.[12, 13]

These results underline the importance of awareness amongst physicians of the preference sensitive nature of the decision to participate in cancer screening programs. Hence, physicians should aim at informing patients about all relevant risks and benefits, and should also provide patients with accurate and individual estimates on their risk of developing the cancer of interest. Personalized risk estimates can be obtained with risk prediction tools, which are widely available for several types of cancer.[15-17] Risk prediction tools that provide accurate individual risk estimates contain important information for both the patient and the physician. A recent study on women's perception and intended behavior regarding epigenetic risk-tailored cancer screening found that women want to obtain information on their individual cancer risk to use this information in the decision on possible participation.[18] Furthermore, physicians can use these risk estimates to support their clinical opinion and aid patients in making an informed decision.

Limitations

The microsimulation model that was used to estimate the cost-effectiveness of prophylactic hysterectomy in Lynch positive women (Chapter 3) used a fixed utility for prophylactic hysterectomy in the base case analysis and varied this in sensitivity analyses. Although these sensitivity analyses provide an estimate of the impact of variation in utilities on the decision to undergo prophylactic hysterectomy, it may not reflect real life individual (dis)utilities. Given the preference sensitive nature of the decision to undergo prophylactic surgery, it might be desirable to adopt these preferences into our model. However, there is currently no methodology available that makes it possible to incorporate this variation in preferences.

Prediction modelling in cancer

Methodological shortcoming of cancer risk prediction models

Cancer risk prediction models are potentially useful tools to estimate the probability of developing cancer or the probability of dying from a specific cancer. These probabilities are estimated for individual patients based on specific characteristics. The number of developed

prediction models is continuously increasing, which makes it difficult to determine their usefulness in clinical practice. We found that only a few models have been developed to predict endometrial cancer in asymptomatic or symptomatic women (Chapter 4). The usefulness of most models is unclear considering methodological shortcomings and lack of external validation. These methodological shortcomings are not specific for risk prediction models for endometrial cancer and are also present in other research areas such as cardiovascular disease and neurology.[19-21] Unfortunately, there are still large numbers of prediction models with methodological shortcomings, which unduly might be used in clinical practice (Table 2).

Model development

Regarding model development, the most important shortcomings were found in the handling of missing data and the selection of predictors. The majority of developed models did not mention the presence of missing data and handling of missing data was performed with suboptimal measures such as creating an indicator variable or single, simple imputation. It has been proven that these methods could lead to biased results.[22] A modern, elegant approach is to impute missing data with multiple imputation methods.[22, 23] Furthermore, transparency on the amount of missing data present in the original data is important to assess the quality of the data and analyses.[24]

Selection of predictors for the final multivariable model was often performed with stringent p-values of 0.05 as stopping rule for selection of predictors in relatively small sample sizes. The use of stringent p-values in combination with small sample sizes will lead to overfitting.[23] The developed model is too specific for the data and will not provide valid predictions for new pations; not for patients from the same setting (internal validity) and not for different settings (external validity).[25] Furthermore, this may lead to too optimistic estimates on the performance of a model in terms of discriminative ability. Hence, the final prediction model may not be useful in new populations as the resulting predictions will be too extreme.

We did not use the recently proposed PROBAST instrument for risk of bias assessment, as this instrument was not available during our study. This instrument contains 4 domains and 20 questions to assess the risk of bias of diagnostic models and prediction models.[26] Although this instrument is more elaborate than our assessment, there is some overlap in topics such as ‘predictors’ and ‘analysis’.

Model validation

For model performance, internal and external validation studies were limited or completely lacking. Internal validation might best be performed with bootstrapping, as this is the most efficient and allows to develop the model on the complete data instead of a part of the data when the split-sample method is used.[27]

External validation is a vital step before implementing prediction models in clinical practice. However, this step is often forgotten or ignored. This may have serious consequences, because results of the original model may not be generalizable and applicable in new patients.[28] External validation can be performed by analyzing the discrimination (ability to distinguish between patients with an without the outcome of interest) and calibration (agreement between observed outcomes and predicted outcomes) of the model in a new population. Afterwards, the original prediction model can be adjusted based on the results of the external validation.[24] Adjustment of existing models is preferred over developing new models, since developing a new model would require to start the process from model development to external validation all over again.[29]

Even if models are developed without methodological shortcomings, this does not automatically lead to models with a high apparent discriminative ability. A possible reason could be that traditional epidemiological predictors do not contain enough information on the outcome of interest. Technological advances have made it possible to obtain information from the (epi)genome, which could be added to existing prediction models and possibly improve their performance.[30-32]

Table 2. Methodological shortcomings in prediction models and possible solutions

Shortcoming	Solution
Model development	
Missing data: no reporting on missing data and use of inefficient handling methods	Report the amount of missing data per variable and use multiple imputation methods to impute the missing data
Selection of predictors for final model: use of stringent p-values (e.g. 0.05) as stopping rule in small sample sizes	Use Akaike's Information Criterion to limit overfitting, penalization, and/or a pre-defined model specification
Model performance	
Internal validation: no internal validation was performed or inefficient methods were used (e.g. split-sample method)	Use bootstrapping to limit overfitting and to correct for optimism
External validation: lack of external validation	Perform external validation before using prediction models in clinical practice, determine model calibration and assess intercept and slope

Application of risk prediction models in clinical practice

In this thesis, several prediction models have been developed and internally validated according to state-of-the-art methodology (table 3).

Chapter 5 showed the development and validation of a model to estimate the probability of regression of large hepatocellular adenomas (benign liver tumors). We found that hepatocellular adenoma diameter at diagnosis, regression-over-time, and the inflammatory

subtype of hepatocellular adenoma and the unclassified subtype (compared to hepatocyte nuclear factor 1α inactivated HCA), were the strongest predictors of regression. The model yielded an internally validated c-index of 0.79.[33]

In Chapter 6 we developed a model to predict the risk of developing keratinocyte carcinoma (type of skin cancer) in patients with actinic keratosis (premalignant lesion of skin cancer). The final model comprised the presence of more than four actinic keratosis, actinic keratosis localization on the upper extremities or elsewhere except the head and coffee consumption. The final model was internally validated with bootstrapping and showed a c-index of 0.6.[34]

A prediction model to identify pancreatic intraductal papillary mucinous neoplasms (neoplastic pancreas cysts) at lowest risk of progression was developed in Chapter 7. Significant predictors of regression were cyst size, cyst multifocality, smoking status, history of acute pancreatitis, and history of extrapancreatic malignancy. After internal validation, the model had a c-index of 0.7.[35]

Although the models that were developed in Chapters 5, 6 and 7 have different outcomes of interest, they also have important similarities. After external validation studies have been performed, all three models might be used to distinguish high-risk patients from low-risk patients to develop a specific type of cancer. This distinction is important because it can help physicians in determining the appropriate treatment strategy and optimal follow-up length. For example, the model that was developed in chapter 5 may aid physicians to identify patients at with benign liver tumors at low risk of progression to liver cancer in order to reserve invasive treatment for high-risk patients. This is beneficial for both the patient and society, as it results in fewer patients receiving unnecessary invasive treatment, which possibly leads to a reduction in health care expenditure. All three models may complement physicians view with objective estimates on individuals' patients risks. However, before implementing these models in practice, they should be externally validated, which makes it possible to determine model calibration. Since our models have only been internally validated, assessment of calibration was considered non-informative. Ideally, external validation should be followed by an impact study. An impact study gives insights into the effects of using the prediction model on patient outcome, cost-effectiveness or other relevant outcome measures.[29] With an impact study, it is also possible to study changes in doctor's behavior. In patients with hepatocellular adenoma, this could mean that an impact study could show the difference in referral practice for invasive treatment between physicians who use the prediction tool and who do not use the tool. An impact analysis would provide insights in all these aspects, which will aid in eventually implementing risk prediction models in clinical practice.

Table 3. Characteristics of prediction models developed in this thesis

	Chapter 5	Chapter 6	Chapter 7
Area of interest	Benign liver tumors	Skin cancer	Pancreas cysts
N (development)	180	1169	875
Validation	Internal validation with bootstrapping	Internal validation with bootstrapping	Internal-external validation
C-statistic (internally validated)	0.79	0.6	0.7

Limitations

A common drawback in the development of prediction models is the unavailability of all clinically relevant predictors. Due to discrepancies in the measurement of certain predictors, we did not have access to all predictors that were planned to be incorporated in our model. This may have influenced the discriminative ability of the models displayed in this thesis. Hence, it is important to give data collection sufficient attention, as the performance of the prediction model is highly dependent on the quality of the data. Despite an extensive search for possible external populations to validate the models for patients with hepatocellular adenoma (Chapter 5), actinic keratosis (Chapter 6) and neoplastic pancreas cysts (Chapter 7), we did not succeed in performing an external validation for these models. As mentioned before, this is a crucial step before implementing a model in clinical practice, since it provides insight into the generalizability to new populations and quantifies the heterogeneity of risk predictors of different populations.

Combining cancer risk prediction and cancer screening

As described before, both better risk prediction and cancer screening can individually aid decision making in cancer since patients might be more informed because of the use of better screening tests and prediction models. Moreover, a combination of cancer risk prediction and cancer screening might enlarge the effectiveness on improving patient outcomes and lower the societal burden of cancer care.

One example of combining risk prediction and cancer screening is the use of a prediction model to identify individuals with Lynch Syndrome amongst patients with colorectal cancer. In theory, it is possible to test every colorectal cancer patient for the presence of Lynch Syndrome. However, this would lead to inefficient use of health care resources (e.g. high rate of negative tests, unnecessary surveillance) and negative patient outcomes (e.g. unnecessary stress of possible Lynch Syndrome diagnosis). Therefore, current guidelines have recommended to use the MMRPredict, MMRPro, and $PREMM_{1,2,6}$ prediction models to distinguish between patients with low and high risk of Lynch Syndrome.[36] Combined with a risk threshold, such as 5% risk, these models can be used to select patients who are likely to have Lynch Syndrome and who will benefit from more frequent diagnostic workup,

prophylactic surgery and other preventive strategies.[36] As a consequence of using these prediction models, only the individuals who benefit from more intensive surveillance and follow-up will receive it, while low-risk individuals remain free from harm.

Another promising effort in combining risk prediction with a nationwide screening program is the FORECEE project. This project aims at implementing an omics based risk prediction for women's cancers (breast, ovarian, endometrial and cervical cancer) into current screening programs with pap smears.[37] Based on analysis of the epigenome, genome, and metagenome, an individual estimate on women's 1-, 2- and 5-year risk will be derived. With these estimates, it is possible to stratify women according to their risk (e.g. low, medium and high-risk women). Each risk group will receive further preventive and screening strategies that are suitable for their risk. The high-risk group will receive more intensive surveillance such as yearly mammography screening, while the low risk group might be advised to undergo mammography once every five year. Hence, this risk stratification results in a lower burden for both individuals and society.

In conclusion, combining risk prediction of cancer with cancer screening is likely to improve improve patient outcomes by reducing harms and improving benefits because of targeted interventions instead of mass population strategies. Furthermore, a combination of prediction and screening will lower health care expenditure as resources are used more efficiently.

Conclusions

Our conclusions on ethics, cost-effectiveness, and prediction modeling of cancer screening and prediction are as follows.

- Patient autonomy should receive additional attention when epigenetic risk-tailored cancer screening is implemented and could be supported by patient decision aids.
- Prophylactic hysterectomy seems a promising strategy to avoid endometrial cancer cases and endometrial cancer related deaths. The optimal age of having prophylactic surgery depends on individuals' preferences.
- There are only a few models available to predict endometrial cancer and most developed models have methodological shortcomings, such as the handling of missing data and the lack of internal and external validation.
- Prediction models should be developed according to proposed guidelines to obtain accurate individual risk estimates that can aid in medical decision making.

Future directions and implications for research or clinical practice

Based on this thesis, recommendations can be made for cancer screening and prediction (Box 3).

In **cancer screening,** further research should focus on incorporating individuals' preferences into microsimulation modelling, to reflect the real life variation in (dis)utility for prophylactic surgery. Furthermore, more insights into the trade-offs that individuals make when weighing benefits and harms of cancer screening programs are necessary to estimate the amount of variation in preferences. If results from analysis of the epigenome are included in screening programs, specific attention should be paid in sustaining patient autonomy.

In **cancer prediction,** more attention should be paid to correct model development and validation. Existing models should be validated and updated, instead of the current trend of developing more and more new models. Impact studies might be useful tools to estimate the effect of the use of prediction models in clinical practice.

A combination of cancer prediction and cancer screening might result in more efficient cancer screening strategies. This could improve patient outcomes and reduce health care expenditure on cancer care.

Box 3 – Recommendations for cancer screening and prediction

Cancer screening

- Acknowledge the importance of patient autonomy in cancer screening programs and support patients by providing all relevant information on their decision, possibly supported by patient decision aids
- Incorporate individuals' utility on hysterectomy and premature menopause to determine the optimal starting age of performing a prophylactic hysterectomy
- Incorporate cancer risk prediction models in cancer screening to improve patient outcomes and lower societal burden of cancer care

Cancer prediction

- Develop prediction models according to proposed guidelines to prevent bias, to limit overfitting, and to improve potential generalizability
- Perform internal and external validation to determine the performance of the prediction model in new patients across a range of settings
- Incorporate new information such as epigenetics in existing prediction models to improve model performance
- Perform an impact study to estimate the effects of using a prediction model in clinical practice

References

1. Vos IML, Schermer MHN, Bolt ILLE. Recent insights into decision-making and their implications for informed consent. Journal of Medical Ethics. 2018;44: 734.
2. Stacey D, Legare F, Lewis K, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. 2017;4: CD001431.
3. Elwyn G, O'Connor A, Stacey D, et al. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. BMJ (Clinical research ed). 2006;333: 417.
4. Adam Husney MD - Family Medicine & Kathleen Romito MD - Family Medicine & E. Gregory Thompson MD - Internal Medicine & Wendy Y. Chen MD MM M-M. Breast Cancer Screening: When Should I Start Having Mammograms? Available from URL: https://www.healthwise.net/ohridecisionaid/Content/StdDocument.aspx?DOCHWID=abh0460 [accessed July 2019, 2019].
5. Holmes-Rovner M, Nelson WL, Pignone M, et al. Are Patient Decision Aids the Best Way to Improve Clinical Decision Making? Report of the IPDAS Symposium. Medical Decision Making. 2007;27: 599-608.
6. Coulter A, Stilwell D, Kryworuchko J, Mullen PD, Ng CJ, van der Weijden T. A systematic development process for patient decision aids. BMC medical informatics and decision making. 2013;13 Suppl 2: S2-S2.
7. Lévesque E, Kirby E, Bolt I, et al. Ethical, Legal, and Regulatory Issues for the Implementation of Omics-Based Risk Prediction of Women's Cancer: Points to Consider. Public Health Genomics. 2018;21: 37-44.
8. Caverly TJ, Cao P, Hayward RA, Meza R. Identifying Patients for Whom Lung Cancer Screening Is Preference-Sensitive. Annals of Internal Medicine. 2018;169: 1-9.
9. Metcalfe KA, Lubinski J, Ghadirian P, et al. Predictors of contralateral prophylactic mastectomy in women with a BRCA1 or BRCA2 mutation: the Hereditary Breast Cancer Clinical Study Group. J Clin Oncol. 2008;26: 1093-1097.
10. Skytte AB, Gerdes AM, Andersen MK, et al. Risk-reducing mastectomy and salpingo-oophorectomy in unaffected BRCA mutation carriers: uptake and timing. Clin Genet. 2010;77: 342-349.
11. Metcalfe KA, Ghadirian P, Rosen B, et al. Variation in rates of uptake of preventive options by Canadian women carrying the BRCA1 or BRCA2 genetic mutation. Open medicine : a peer-reviewed, independent, open-access journal. 2007;1: e92-e98.
12. Tiller K, Meiser B, Butow P, et al. Psychological impact of prophylactic oophorectomy in women at increased risk of developing ovarian cancer: a prospective study. Gynecol Oncol. 2002;86: 212-219.
13. Beattie MS, Crawford B, Lin F, Vittinghoff E, Ziegler J. Uptake, time course, and predictors of risk-reducing surgeries in BRCA carriers. Genetic testing and molecular biomarkers. 2009;13: 51-56.
14. Hatcher MB, Fallowfield L, A'Hern R. The psychosocial impact of bilateral prophylactic mastectomy: prospective study using questionnaires and semistructured interviews. BMJ (Clinical research ed). 2001;322: 76.
15. Fortner RT, Husing A, Kuhn T, et al. Endometrial cancer risk prediction including serum-based biomarkers: results from the EPIC cohort. Int J Cancer. 2017;140: 1317-1323.

16. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. Breast Cancer Res Treat. 2012;132: 365-377.
17. Pfeiffer RM, Park Y, Kreimer AR, et al. Risk prediction for breast, endometrial, and ovarian cancer in white women aged 50 y or older: derivation and validation from population-based cohort studies. PLoS medicine. 2013;10: e1001492.
18. Wegwarth O, Pashayan N, Widschwendter M, Rebitschek FG. Women's perception, attitudes, and intended behavior towards predictive epigenetic risk testing for female cancers in 5 European countries: a cross-sectional online survey. BMC public health. 2019;19: 667-667.
19. Damen JA, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ (Clinical research ed). 2016;353: i2416.
20. Mushkudiani NA, Hukkelhoven CW, Hernandez AV, et al. A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes. J Clin Epidemiol. 2008;61: 331-343.
21. Dijkland SA, Foks KA, Polinder S, et al. Prognosis in Moderate and Severe Traumatic Brain Injury: A Systematic Review of Contemporary Models and Validation Studies. J Neurotrauma. 2020;37: 1-13.
22. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol. 2006;59: 1087-1091.
23. Steyerberg EW. Clinical Prediction Models. Springer-Verlag New York, 2009.
24. Vergouwe Y, Royston P, Moons KGM, Altman DG. Development and validation of a prediction model with missing predictor data: a practical approach. Journal of Clinical Epidemiology. 2010;63: 205-214.
25. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med. 1999;130: 515-524.
26. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Annals of Internal Medicine. 2019;170: 51-58.
27. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart. 2012;98: 683-690.
28. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. Bmj. 2009;338: b605.
29. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ (Clinical research ed). 2009;338: b606.
30. Wahl S, Drong A, Lehne B, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2017;541: 81-86.
31. Bojesen SE, Timpson N, Relton C, Davey Smith G, Nordestgaard BG. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. Thorax. 2017;72: 646-653.
32. Widschwendter M, Apostolidou S, Raum E, et al. Epigenotyping in peripheral blood cell DNA and breast cancer risk: a proof of principle study. PLoS One. 2008;3: e2656.
33. Klompenhouwer AJ, Alblas M, van Rosmalen BV, et al. Development and Validation of a Model to Predict Regression of Large Size Hepatocellular Adenoma. Am J Gastroenterol. 2019;114: 1292-1298.

34. Tokez S, Alblas M, Nijsten T, Pardo LM, Wakkee M. Predicting keratinocyte carcinoma in patients with actinic keratosis: development and internal validation of a multivariable risk-prediction model. Br J Dermatol. 2020;183: 495-502.
35. Overbeek KA, Alblas M, Gausman V, et al. Development of a stratification tool to identify pancreatic intraductal papillary mucinous neoplasms at lowest risk of progression. Alimentary pharmacology & therapeutics. 2019;50: 789-799.
36. Kastrinos F, Ojha RP, Leenen C, et al. Comparison of Prediction Models for Lynch Syndrome Among Individuals With Colorectal Cancer. LID - 10.1093/jnci/djv308 [doi] LID - djv308 [pii].
37. FORECEE Female Caner Prediction Using Cervical Omics to Individualise Screening and Prevention. Available from URL: http://www.forecee.eu/ [accessed November 20 2017, 2017].

Appendices

Summary

The general introduction, **Chapter 1**, provides an overview of the background and aims addressed in this thesis. The global cancer burden has increased over the past decades and cancer incidence and mortality are expected to spike even further in the future. Hence, specific actions in the area of cancer screening and risk prediction are necessary to change cancer's course. Cancer screening entails testing of asymptomatic individuals for abnormalities that indicate the presence of pre-cancer or (early stage) cancer. Because of the timing of screening (before individuals become symptomatic), it often results in better patient outcomes in terms of mortality, morbidity and treatment possibilities. In the field of cancer, risk prediction models can be used to estimate the probability of cancer, or to estimate the probability of progression from a premalignant lesion to clinical cancer. Cancer screening and prediction can reinforce each other, as risk prediction allows to target cancer screening to those individuals who will benefit most from it. The overall aim of this thesis was to evaluate ethics, cost-effectiveness, and prediction modeling of cancer screening and treatment. Specific research questions were:

1. How can we promote autonomous choices in informed consent procedures regarding participation in epigenetic risk-tailored cancer screening?

2. What is the cost-effectiveness of offering prophylactic hysterectomy to first-degree relatives with Lynch Syndrome of patients with colorectal cancer?

3. What is the quality of risk prediction models for endometrial cancer in the general population?

4. How can we obtain accurate and valid individual estimates of progression from premalignant to malignant cancer based on prediction models to guide decision making?

Screening

In **Chapter 2**, we found that new informed consent models, such as the tiered-layered-staged informed consent, are highly suitable for use in epigenetic cancer screening programs. Furthermore, we found that patient decision aids (PDAs) might support patients in making an autonomous choice regarding participation in a cancer screening program. We proposed a framework that guides the development of Patient Decision Aids (PDAs) to support informed consent in the specific context of epigenetic cancer screening programs.

In **Chapter 3**, we studied the cost-effectiveness of offering prophylactic hysterectomy in first-degree female relatives with Lynch Syndrome of probands with colorectal cancer. We found that prophylactic hysterectomy is most cost-effective when offered to women aged 40-80 years. However, the decision to undergo prophylactic surgery remains a preference-sensitive

decision, and depends on individuals' perception and disutility of prophylactic surgery and its consequence such as premature menopause.

Prediction

In **Chapter 4**, we found that only a few models have been developed to predict endometrial cancer in asymptomatic or symptomatic women. The usefulness of most models is unclear considering methodological shortcomings and lack of external validation. Risk prediction models should be externally validated before their use in clinical practice to prevent that resulting risk predictions that are too high or too low form the base of treatment decisions.

In this thesis, several prediction models have been developed and internally validated according to state-of-the-art methodology.

Chapter 5 showed the development and validation of a model to estimate the probability of regression of large hepatocellular adenomas (benign liver tumors). We found that hepatocellular adenoma diameter at diagnosis, regression-over-time, and the inflammatory subtype of hepatocellular adenoma and the unclassified subtype (compared to hepatocyte nuclear factor 1α inactivated HCA), were the strongest predictors of regression. The model yielded an internally validated c-index of 0.79.

In **Chapter 6**, we developed a model to predict the risk of developing keratinocyte carcinoma (type of skin cancer) in patients with actinic keratosis (premalignant lesion of skin cancer). The final model comprised the presence of more than four actinic keratosis, actinic keratosis localization on the upper extremities or elsewhere except the head and coffee consumption. The final model was internally validated with bootstrapping and showed a c-index of 0.6.

A prediction model to identify pancreatic intraductal papillary mucinous neoplasms (neoplastic pancreas cysts) at lowest risk of progression was developed in **Chapter 7**. Significant predictors of regression were cyst size, cyst multifocality, smoking status, history of acute pancreatitis, and history of extrapancreatic malignancy. After internal validation, the model had a c-index of 0.7.

Chapter 8 discussed the results and implications of this thesis. The aim of this thesis was to evaluate ethics, cost-effectiveness, and prediction modeling of cancer screening and treatment.

We found that patient autonomy should receive additional attention when epigenetic risk-tailored cancer screening is implemented and could be supported by patient decision aids. Next, we found that prophylactic hysterectomy seems a promising strategy to avoid endometrial cancer cases and endometrial cancer related deaths. The optimal age of having

prophylactic surgery depends on individuals' preferences. Furthermore, we showed that only a few models are available to predict endometrial cancer and most developed models have methodological shortcomings, such as the handling of missing data and the lack of internal and external validation. Prediction models should be developed according to proposed guidelines to obtain accurate individual risk estimates that can aid in medical decision making. A combination of cancer prediction and cancer screening might result in more efficient cancer screening strategies. This could improve patient outcomes and reduce health care expenditure on cancer care. Based on our findings, we recommend the following in the field of cancer screening and prediction:

Cancer screening

- Acknowledge the importance of patient autonomy in cancer screening programs and support patients by providing all relevant information on their decision, possibly supported by patient decision aids.
- Prophylactic hysterectomy seems a promising strategy to avoid endometrial related cancer deaths. Incorporate individuals' utility on prophylactic hysterectomy and premature menopause to determine the optimal starting age.
- Incorporate cancer risk prediction models in cancer screening to improve patient outcomes and lower societal burden of cancer care.

Cancer prediction

- Develop prediction models according to proposed guidelines to prevent bias, to limit overfitting, and to improve generalizability.
- Perform internal and external validation to determine the performance of the prediction model in new patients.
- Incorporate new information such as epigenetics in existing prediction models to improve model performance.
- Use an impact study to estimate the effects of using a prediction model in clinical practice.

Samenvatting

De algemene introductie, **Hoofdstuk 1,** geeft een overzicht van de achtergrond en de doelen zoals beschreven in dit proefschrift. De wereldwijde, maatschappelijke last als gevolg van kanker is de afgelopen eeuwen gestegen en de verwachting is dat de incidentie en mortaliteit van kanker nog verder door zal stijgen in de toekomst. Er zijn daarom specifieke interventies nodig op het gebied van kanker screening en predictie om een koerswijziging te realiseren. Kanker screening omvat het testen van asymptomatische individuen op afwijkingen die kunnen duiden op de aanwezigheid van (een voorstadium) van kanker. Kanker screening leidt vaak tot verbeterde uitkomsten op het gebied van mortaliteit, morbiditeit en behandelingsopties vanwege de timing van screening (voordat individuen symptomatisch worden). Op het gebied van kanker kunnen risico predictiemodellen worden gebruikt om een inschatting te maken van de kans op kanker of de kans op progressie van een premaligne stadia tot klinische kanker.

Kanker screening en predictie kunnen elkaar versterken doordat risico predictie het mogelijk maakt om kanker screening specifiek te richten op de individuen die het meeste profijt zullen hebben van de screening. Het overkoepelende doel van dit proefschrift was om de ethiek, kosteneffectiviteit en predictie modellering van kanker screening en behandeling te evalueren. Specifieke onderzoeksvragen waren:

1. Hoe kunnen we autonome keuzes promoten in informed consent procedures rondom deelname in epigenetische cancer screening toegespitst op specifieke risicogroepen?

2. Wat is de kosteneffectiviteit van het aanbieden van profylactische hysterectomie aan eerstegraads familieleden met Lynch Syndroom van patiënten met colorectale kanker?

3. Wat is de kwaliteit van risico predictiemodellen voor endometriumkanker in de algemene populatie?

4. Hoe kunnen we accurate en valide individuele schattingen van het risico op progressie van premaligne stadia naar klinische kanker verkrijgen, gebaseerd op predictie modellen om beslisvorming te ondersteunen?

Screening
In **Hoofdstuk 2,** vonden we dat nieuwe informed consent modellen zoals het 'tiered-layered-staged' informed consent geschikt zijn voor gebruik in epigenetische kanker screening programma's. Daarnaast vonden we dat keuzehulpen patiënten kunnen ondersteunen in het maken van een autonome keuze over deelname in een kanker screening programma. We ontwikkelden een raamwerk voor de ontwikkeling van keuzehulpen om informed consent te ondersteunen in de specifieke context van epigenetische kanker screening programma's.

In **Hoofdstuk 3**, onderzochten we de kosteneffectiviteit van het aanbieden van profylactische hysterectomie aan eerstegraads familieleden met Lynch Syndroom van patiënten met colorectale kanker. We vonden dat profylactische hysterectomie het meest kosteneffectief is als dit wordt aangeboden aan vrouwen in de leeftijd van 40 tot 80 jaar. De beslissing om een profylactische operatie te ondergaan blijft een individuele beslissing en is afhankelijk van de individuele perceptie en disutiliteit van profylactische hysterectomie en de bijbehorende consequenties zoals vroegtijdige menopauze.

Prediction

In **Hoofdstuk 4,** vonden we dat er maar een beperkt aantal modellen is ontwikkeld om het risico op endometrium kanker te voorspellen in asymptomatische of symptomatische vrouwen. De bruikbaarheid van de meeste modellen is onduidelijk vanwege methodologische tekortkomingen en het gebrek aan externe validatie. Risico predictie modellen dienen extern gevalideerd te worden voordat ze gebruikt worden in de klinische praktijk om te voorkomen dat foutieve voorspellingen (te hoog of te laag) worden gebruikt als basis voor beslissingen over behandelingen.

In dit proefschrift zijn verschillende predictie modellen ontwikkeld en intern gevalideerd volgens hoogstaande methodologie.

Hoofdstuk 5 liet de ontwikkeling en validatie zien van een model om de kans te berekenen van regressie van grote hepatocellulaire adenomen (goedaardige lever tumoren). We vonden dat de diameter van het adenoom bij diagnose, de regressie over tijd en het subtype van het adenoom de sterkste voorspellers waren van regressie. Het model gaf een intern gevalideerde c-index van 0.79.

In **Hoofdstuk 6,** ontwikkelden we een model om het risico op het ontwikkelen van keratinocyt carcinoma (huidkanker) bij patiënten met actinische keratose (een premaligne laesie van huidkanker). Het ontwikkelde model bevatte de voorspellers: aanwezigheid van meer dan vier premaligne laesies, lokalisatie van de premaligne laesie op de bovenste extremiteiten of ergens anders behalve op het hoofd en consumptie van koffie. Het model werd intern gevalideerd met bootstrapping en gaf een c-index van 0.6.

Een predictiemodel om neoplastische pancreas cysten te identificeren werd ontwikkeld in **Hoofdstuk 7**. De grootte van de cyste, multivocaliteit, roken, voorgeschiedenis met acute pancreatitis en voorgeschiedenis met ander maligniteiten waren significante voorspellers van regressie. Na interne validatie had het model een c-index van 0.7.

In **Hoofdstuk 8** werden de resultaten en de implicaties van dit proefschrift besproken. Het doel van dit proefschrift was om de ethiek, kosteneffectiviteit en predictie modellering van kanker screening en behandeling te evalueren.

We vonden dat de autonomie van patiënten extra aandacht zou moeten krijgen en kan worden ondersteund met keuzehulpen als epigenetische kanker screening programma's worden geïmplementeerd. Daarnaast vonden we dat profylactische hysterectomie een veelbelovende strategie lijkt om endometrium kanker en overlijden door endometrium kanker te voorkomen. De optimale leeftijd voor het ondergaan van profylactische hysterectomie hangt af van individuele voorkeuren. Vervolgens hebben we laten zien dat er slecht een aantal modellen beschikbaar zijn om het risico op endometrium kanker te voorspellen en dat de meeste modellen methodologische tekortkomingen hebben, zoals het omgaan met ontbrekende data en het ontbreken van interne en externe validatie. Predictie modellen zouden ontwikkeld moeten worden volgens voorgestelde richtlijnen om accurate individuele risico schatting te verkrijgen die bruikbaar zijn om medische besliskunde te ondersteunen.

Een combinatie van kanker predictie en kanker screening zou kunnen resulteren in efficiëntere kanker screening strategieën. Dit kan leiden tot een verbetering in patiëntuitkomsten en een reductie van kosten voor gezondheidszorg op het gebied van kanker. Gebaseerd op onze bevindingen raden we het volgende aan op het gebied van kanker screening en predictie:

Kanker screening

- Erken het belang van autonomie van patiënten in kanker screening programma's en ondersteun patiënten door alle relevante informatie over beschikbaar te stellen, eventueel ondersteunt met keuzehulpen.
- Profylactische hysterectomie lijkt een veelbelovende strategie om overlijden door endometrium kanker te voorkomen. Bij het bepalen van de optimale startleeftijd moeten individuele voorkeuren over profylactische hysterectomie en vroegtijdige menopauze worden meegenomen.
- Predictie modellen om het risico op kanker te voorspellen zouden moeten worden meegenomen in kanker screening om uitkomsten voor de patiënten te verbeteren en de maatschappelijke last van kanker te verlagen.

Kanker predictie

- Ontwikkel predictiemodellen volgens voorgestelde richtlijnen om bias te voorkomen, om overfitting te beperken en om generaliseerbaarheid te vergroten.
- Verricht interne en extere validatie om de prestatie van het predictiemodel in nieuwe patiënten vast te stellen

- Verwerk nieuwe informatie zoals epigenetica in bestaande predictiemodellen om de prestaties van de modellen te verbeteren
- Gebruik een impact studie om het effect te schatten van het gebruik van een predictiemodel in de klinische praktijk

List of publications

Alblas M, Velt KB, Pashayan N, Widschwendter M, Steyerberg EW, Vergouwe Y.Prediction models for endometrial cancer for the general population or symptomatic women: A systematic review. Crit Rev Oncol Hematol. 2018 Jun;126:92-99.

Salampessy BH, **Alblas MM**, Portrait FRM, Koolman X, van der Hijden EJE The effect of cost-sharing design characteristics on use of health care recommended by the treating physician; a discrete choice experiment. BMC Health Serv Res. 2018 Oct 20;18(1):797.

Alblas M, Schermer M, Vergouwe Y, Bolt I. Autonomy Challenges in Epigenetic Risk-Stratified Cancer Screening: How Can Patient Decision Aids Support Informed Consent? J Pers Med. 2019 Feb 18;9(1)

Klompenhouwer AJ, **Alblas M**, Vivica van Rosmalen B, Haring MPD, Venema E, Doukas M, Thomeer MGJ, Takkenberg RB, Verheij J, de Meijer VE, van Gulik TM, Lingsma HF, de Man RA, Ijzermans JNM. Development and Validation of a Model to Predict Regression of Large Size Hepatocellular Adenoma. Am J Gastroenterol. 2019 Mar 27.

Kasper A. Overbeek, **Maaike Alblas**, Valerie Gausman, Pujan Kandel, Adam B. Schweber, Christian Brooks, Priscilla A. Van Riet, Michael B. Wallace, Tamas A. Gonda, Djuna L. Cahen, Marco J. Bruno. Development of a stratification tool to identify pancreatic intraductal papillary mucinous neoplasms at lowest risk of progression. Aliment Pharmacol Ther . 2019 Oct;50(7):789-799

S. Tokez, **M. Alblas**, T. Nijsten, L.M. Pardo Cortes, H.F. Lingsma, M. Wakkee. Risk of a primary keratinocyte carcinoma in patients with actinic keratosis: a prognostic model. Br J Dermatol. 2019 Dec 19.

R I Neuman, **M M Alblas van der Meer**, L Saleh, S A A van den Berg, A H van den Meiracker, A H J Danser, W Visser. Copeptin and mid-regional pro-ANP in women with suspected or confirmed pre-eclampsia: comparison with the sFlt-1/PlGF ratio. Ultrasound Obstet Gynecol. 2020 Jan 24

L. Saleh, **M. Alblas**, D. Nieboer, R. Neuman, Y. Vergouwe, I. Brussé, J. J. Duvekot, E.W. Steyerberg, H.J. Versendaal, J.A.H. Danser, A.H. van den Meiracker, K. Verdonk, W. Visser. Prediction of pre-eclampsia-related complications in women with suspected/confirmed pre-eclampsia: development and internal validation of a clinical prediction model. Ultrasound Obstet Gynecol. 2020 Oct 8.

Neuman RI, **Alblas van der Meer MM**, Nieboer D, Saleh L, Verdonk K, Kalra B, Kumar A, Alpadi K, van den Meiracker AH, Visser W, Danser AHJ.
PAPP-A2 and Inhibin A as Novel Predictors for Pregnancy Complications in Women With Suspected or Confirmed Preeclampsia. J Am Heart Assoc. 2020 Oct 20;9(19):e018219.

Maaike Alblas, Elisabeth F.P. Peterse, Mengmeng Du, Ann G. Zauber, Ewout W. Steyerberg, Nikki van Leeuwen, Iris Lansdorp-Vogelaar. Cost-effectiveness of prophylactic hysterectomy in first-degree female relatives with Lynch syndrome of patients diagnosed with colorectal cancer in the United States: a microsimulation study. Cancer Med. 2021 Sep 12

PhD portfolio

Name PhD Candidate: Maaike Alblas
Erasmus MC Department: Public Health
Phd Period: Aug 2016 - Aug 2019

Promotor: Prof. dr. E.W. Steyerberg
Copromotor: dr. N. van Leeuwen

1. PhD training	Year	Workload (ECTS)
Research skills		
Introduction R	2016	0.3
Systematical literature retrieval	2016	0.6
Research Integrity Course	2017	0.3
CPO course – Center for Patient Oriented Research	2017	0.3
Research courses		
Advanced Topics in Decision-making in Medicine, NIHES Rotterdam	2018	2.4
Survival Analysis, NIHES Rotterdam	2016	1.4
Markers and Prediction Research, NIHES Rotterdam	2016	0.7
Competing Risks, KU Leuven	2016	1.0
Advanced analysis of prognosis studies, NIHES Rotterdam	2017	0.9
Presentations at national and international conferences		
A new framework for decision aids to support informed consent and ensure autonomous choices with an application in female cancer screening SMDM 2017, Pittsburgh – United States (poster)	2017	1.5
Cost-effectiveness of Prophylactic Hysterectomy in First-Degree Female Relatives with Lynch Syndrome SMDM 2018, Leiden – the Netherlands (poster)	2018	1.0
Autonomy challenges in (epi)genetic risk-stratified cancer screening. Consortium meeting FORECEE, Rotterdam – the Netherlands	2017	1.5
Research meeting CMB/MBL, department of Public Health, Erasmus MC Rotterdam – the Netherlands	2017	1.0
Cost-effectiveness of Prophylactic Hysterectomy in First-Degree Female Relatives with Lynch Syndrome Club meth, Erasmus MC Rotterdam – the Netherlands	2018	1.0
National and international conferences		
Consortium meeting FORECEE Milan – Italy	2016	1.0
Consortium meeting FORECEE Berlin – Germany	2017	1.0
Consortium meeting FORECEE London – United Kingdom	2017	1.0
Teleconferences FORECEE	2016-2019	2.0

Seminars and workshops		
Research seminars, dep. Public Health, Erasmus MC Rotterdam	2016-2019	3.0
CMB/Club Meth, Erasmus MC Rotterdam	2016-2019	1.0
Erasmus MC PhD Day	2017	0.3
2. Teaching activities		
Supervision medical students theme 3.0.4 (community project)	2016-2019	1.5
Supervision PhD Candidate (Colombia)	2017-2018	3.0
Consultations for clinical departments Erasmus MC Rotterdam	2016-2019	6.0

Dankwoord

Coming together is a beginning.
Keeping together is progress.
Working together is success.

- Henry Ford

Het moment waar ik stiekem al jaren naar heb uitgekeken: het schrijven van mijn dankwoord. Dit betekent namelijk dat mijn proefschrift officieel is afgerond, een grote mijlpaal. Bij het starten van mijn PhD in 2016 had ik niet kunnen bedenken aan welk avontuur ik begon. In de afgelopen jaren heb ik aan mooie projecten mogen werken met inspirerende mensen en heb ik de kans gekregen om te groeien op wetenschappelijk en persoonlijk gebied. Dit proefschrift is het resultaat van een intensieve en unieke periode in mijn leven, waaraan veel mensen op verschillende manieren hebben bijgedragen.

Graag wil ik iedereen bedanken voor hun inhoudelijke bijdrage aan dit proefschrift, de volgende personen in het bijzonder.

Ewout, ik bewonder je uitzonderlijke kritische en scherpe blik en blijf me verbazen over de snelheid waarmee je naar al mijn manuscripten en bijbehorende vragen hebt gekeken. Bedankt voor je vertrouwen en de vrijheid die je mij hebt gegeven in de keuzes rondom de inhoud van dit proefschrift.

Yvonne, onze wegen scheidden zich eerder dan ik had verwacht. Desondanks heb je een waardevolle bijdrage geleverd aan de start van mijn PhD traject en wil ik je bedanken voor de wekelijkse sessies waarin je geduldig al mijn vragen als startende onderzoeker bleef beantwoorden.

Nikki, waar moet ik beginnen? Als co-promotor heb jij mijn PhD traject tot een succes gemaakt. Ik heb genoten van onze wekelijkse meetings die vaak over inhoudelijke zaken gingen, maar ook net zo vaak over andere belangrijke dingen in het leven. Ondanks dat je bij een aantal artikelen niet direct betrokken was als co-auteur was je altijd geïnteresseerd in de onderwerpen en kon ik bij je terecht met al mijn vragen en frustraties. Bedankt voor je oprechte betrokkenheid en eindeloze steun, ook in de periodes waarin het leven soms iets moeilijker was. Je hebt mij laten zien dat loslaten soms best fijn kan zijn en dat fouten maken menselijk is. Dankzij jou ben ik gegroeid als wetenschapper en als mens.

Hester, ondanks dat mijn projecten zich veelal buiten de scope van jouw projecten afspeelden was je altijd bereikbaar voor vragen of overleg. Je hebt me laten zien dat het niet altijd nodig is om een vastomlijnd plan te hebben, dank hiervoor.

Daan, dank voor je bereidheid om al mijn vragen over de statistiek in mijn manuscripten te beantwoorden. Zonder jouw hulp waren mijn foutmeldingen in R nog steeds niet opgelost en was dit proefschrift er nooit gekomen.

Daarnaast wil ik alle CMB-ers bedanken voor hun betrokkenheid en gezelligheid. Ontzettend fijn om altijd de mogelijkheid te hebben om koffie te drinken en te sparren over inhoudelijke onderwerpen. In het bijzonder ook dank aan de ondersteunende diensten, zoals ICT en het secretariaat. Bedankt voor jullie geduld en hulp bij alle vragen en problemen die naar boven kwamen tijdens de afgelopen jaren.

Verder wil ik prof. dr. Kees Verhoef, prof. dr. Marjanka Schmidt en dr. Iris Lansdorp-Vogelaar bedanken voor het beoordelen van dit proefschrift. De overige commissieleden wil ik hartelijk bedanken voor het lezen van mijn proefschrift en het deelnemen aan de oppositie.

Diverse hoofdstukken in dit proefschrift zijn het resultaat van mooie samenwerkingen met andere (klinische) afdelingen van het Erasmus MC en daarbuiten.
Iris, bedankt voor de kans die je me hebt gegeven om te werken aan artikel waarbij we gebruik hebben gemaakt van een MISCAN model. Een compleet nieuwe techniek en manier van denken, ontzettend leerzaam en interessant. Dank voor je kritisch blik en opbouwende feedback op de verschillende versies van het manuscript.
Elleke, zonder jouw hulp was ons artikel geen succes geworden. Bedankt dat ik met al mijn vragen bij je terecht kon. Je gaf me het gevoel dat ik alle vragen die ik had gewoon kon stellen, een ontzettend waardevolle eigenschap.
Ineke en Maartje, bedankt dat jullie mij mee hebben genomen in de wondere wereld van de medische ethiek. Een ontzettend interessant vakgebied waarin veel complexe vraagstukken spelen. Dank voor jullie vertrouwen en waardevolle input in het manuscript over keuzehulpen.
Kasper, Julia, Selin, Langeza en Rugina, bedankt voor de fijne samenwerking. Ik waardeer het geduld waarmee jullie mij de inhoudelijke, klinische achtergrond van onze projecten wilden uitleggen en kijk met veel plezier terug op onze samenwerkingen. Mooi om te zien hoe goed het uitpakt als binnen samenwerkingen iedereen een bijdrage vanuit zijn expertise kan leveren.
Further, I would like to thank all members of the FORECEE consortium for giving me the opportunity to be part of this inspiring international group of researchers, which resulted in two manuscripts that are part of this thesis.

Naast de mensen die een inhoudelijke bijdrage hebben geleverd aan dit proefschrift hebben mijn vrienden en familie mij de afgelopen jaren ontzettend gesteund. Er zijn een aantal mensen die ik in het bijzonder wil bedanken.
Allereerst mijn allerliefste paranimfen, Simone, Kelly en Vicky. Ondanks dat er formeel maar twee paranimfen mogen worden aangewezen zijn jullie voor mij alle drie mijn paranimfen. Ik kijk terug op een onvergetelijke tijd, ooit begonnen in kamer 2424. Onze vele etentjes met sushi, gezellige avonden thuis, tripjes van Zeeland tot aan New York, het zijn allemaal momenten om nooit te vergeten. Ik weet niet wat ik zonder jullie had gemoeten, ook toen ik onverhoopt jullie kennis als (bijna)-neurologen hard nodig had. Ondanks dat we allemaal onze eigen wegen zijn gegaan weten we van elkaar dat we altijd bij elkaar terecht kunnen. Bedankt voor wie jullie zijn, ik kijk uit naar alle mooie momenten die nog gaan komen.
Esther, bedankt voor al je steun en interesse in de afgelopen jaren. Het is ontzettend waardevol om te weten dat ik altijd op jou kan terugvallen. We leerden elkaar kennen als studenten in Groningen en inmiddels zijn we allebei moeder van prachtige kindjes, een heel ander leven met nieuwe uitdagingen en hoogtepunten. Bedankt dat ik altijd bij je terecht kan om mijn hart te luchten en dat je me af en toe een spiegel voorhoudt.

Hannah, we kunnen maanden geen contact hebben om vervolgens drie uur met elkaar te bellen. Bij jou voelt het altijd vertrouwd en zit het altijd goed. Onze reis naar Nieuw-Zeeland was een trip om nooit te vergeten, de herinneringen zijn me ontzettend dierbaar. We hebben allebei een heel ander leven maar tegelijkertijd hebben we ontzettend veel overeenkomsten. Ik bewonder je kracht en doorzettingsvermogen, je hebt de afgelopen jaren laten zien dat je een ontzettend krachtige vrouw bent.
Huiskring – Cornelie, Marien, Adinda, Walter, Tijmen, Esther en Eveline. Ontzettend veel dank voor de gezellige en inspirerende avonden. Jullie hebben altijd interesse gehouden in de vorderingen van mijn proefschrift, ondanks dat het lang(er) duurde. Ik geniet van jullie verschillende karakters en kan van een ieder van jullie iets leren. Bedankt dat jullie er altijd zijn. Ik kijk uit naar nog vele inspirerende avonden en hopelijk kunnen we op korte termijn ons eerste huiskring-weekend organiseren.
Lieve collega's uit Tiel, bedankt voor jullie eindeloze interesse in mijn onderzoek. Als ik het even niet meer zag zitten dan was er altijd wel iemand om me een hart onder de riem te steken en me aan te moedigen om door te gaan. Ik kijk er naar uit om samen met jullie mooie stappen te zetten in 'ons' ziekenhuis.
Lotte, je mag niet ontbreken in mijn dankwoord. Ondanks dat we nog geen lange geschiedenis hebben voelt het alsof we elkaar al jaren kennen. We hebben ontzettend veel overeenkomsten en ik ben heel dankbaar dat we elkaar hebben ontmoet. Bedankt dat je mij tijdens de laatste loodjes van het maken van dit proefschrift hebt bijgestaan. Fijn om altijd bij je terecht te kunnen voor een goed gesprek onder het genot van een kop thee en een bastognekoek.

Mijn lieve familie en schoonfamilie, bedankt voor jullie eindeloze steun en interesse in mijn proefschrift. Lieve heit en mem, bedankt dat jullie altijd voor me klaar staan. Jullie bleven de afgelopen jaren geïnteresseerd in mijn proefschrift. Heel erg fijn dat jullie regelmatig naar Gorinchem kwamen als het door tijdgebrek mij niet lukte om naar Friesland te komen. Bedankt dat jullie altijd achter mij hebben gestaan en de keuzes die ik heb gemaakt hebben ondersteund.
Gerrit en Hanneke, bedankt dat jullie mij zo liefdevol hebben opgenomen in jullie gezin. Ik heb me altijd ontzettend welkom en thuis bij jullie gevoeld. De afgelopen jaren hebben jullie mij op ontelbare manieren geholpen en hier ben ik jullie erg dankbaar voor.
Barbera, Pieter, Heleen, Jochem, Maarten en Nick, bedankt voor alle gezelligheid van de afgelopen jaren. Fijn om te weten dat we altijd bij elkaar terecht kunnen, op welke manier dan ook.
Petra, lief zusje. Wat zie ik veel van mezelf in jou terug en wat maak je mij een trotse grote zus. Je hebt ontzettend veel doorzettingsvermogen en een groot verantwoordelijkheidsgevoel, dit heeft je al ver gebracht. Ik bewonder hoe je alle ballen hoog houdt en de afgelopen jaren je opleidingen, werk en sociale leven hebt weten te combineren met de zorg voor Amora

Jane. Ondanks dat we niet bij elkaar om de hoek wonen voelt het alsof je altijd dichtbij bent, dankjewel hiervoor.

Lieve Jos, zoals je inmiddels weet eindigt het dankwoord met de belangrijkste persoon. Zonder jou zou dit proefschrift er nooit zijn gekomen. Je bent mijn grootste supporter, je helpt me om te relativeren en te vertrouwen op mezelf. Ik kan niet zonder je en kijk uit naar onze toekomst. Wij samen voor altijd, ik hou van je.

About the author

Maaike Alblas was born in Harlingen, the Netherlands on July 9th, 1991. After finishing secondary school at Simon Vestdijk Harlingen, she moved to Groningen to study nursing. Maaike was granted access to follow a special program that consisted of a combination of studying and working as a student-nurse at the University Medical Center Groningen. In 2014, she obtained her Bachelor of Nursing and started a premaster Health Sciences at the Vrije Universiteit Amsterdam. In 2015, she started the master Health Sciences with specialization Health Policy. She conducted her master research at Innopay and investigated the effect of cost-sharing design characteristics on the use of health care recommended by the treating physician with a discrete choice experiment, which resulted in a publication.

In 2016, she started as a PhD candidate at the Department of Public Health of the Erasmus MC under supervision of dr. Yvonne Vergouwe, dr. Nikki van Leeuwen and prof. dr. Ewout Steyerberg. She worked on different international research projects, often in collaboration with clinical departments of the Erasmus MC. She gained experience with different statistical methods, mostly to develop prediction models to predict the risk of developing certain types of cancer. In 2019, she started working at Ziekenhuis Rivierenland Tiel as advisor on Hospital Capacity Management. She lives in Gorinchem with her husband Jos and daughters Linde and Jula.