Participation, risk-selection and effect on mortality in the Danish screening program for colorectal cancer

PhD thesis

Mette Kielsholm Thomsen

September 2022

Department of Clinical Epidemiology

Health

AARHUS UNIVERSITY

"As such, screening is a commendable method to reduce the burden of disease. However, population screening targets a predominantly healthy population, and should therefore only be conducted after a careful consideration of both harms and benefits"

European Commission: "Introduction: Guiding Principles" p. 3 in European Guidelines for Quality Assurance in Colorectal Cancer Screening and Diagnosis, 2010.

Supervisors

Ellen M. Mikkelsen, MPH, PhD (main supervisor)

Associate Professor, Department of Clinical Epidemiology, Aarhus University and Aarhus University Hospital, Denmark

Rune Erichsen, MD, PhD

Physician, Department of Surgery, Randers Regional Hospital, Denmark
Associate Professor, Department of Clinical Epidemiology, Aarhus University and Aarhus University Hospital, Denmark

Lars Pedersen, MSc, PhD

Professor, Department of Clinical Epidemiology, Aarhus University and Aarhus University Hospital, Denmark

Henrik Toft Sørensen, MD, PhD, DMSc, DSc

Clinical professor and chair, Department of Clinical Epidemiology, Aarhus University and Aarhus University Hospital, Denmark

Assessment Committee

Berit Andersen, MD, PhD (Chair)

Professor and head of department, Department of Public Health Programs, Department of Clinical Medicine, Aarhus University and Middle Jutland Region, Denmark

Henrik Bøggild, MD, PhD

Associate professor, Department of Health Science and Technology, Aalborg University, Denmark

Mette Kalager, MD, PhD

Professor, Institute of Health and Society, University of Oslo and Oslo University Hospital, Norway

Preface

This thesis concludes my integrated PhD (combined master's and PhD degree), conducted at the Department of Clinical Epidemiology, the Department of Clinical Medicine, Aarhus University, Denmark. It has entailed two years of combined Master's and PhD degree work and two years of sole PhD project work.

My PhD project investigates various aspects of the Danish colorectal cancer (CRC) screening program. I have been involved in research related to quality assurance of the colorectal cancer screening program since it was launched. Before working on this thesis, I contributed to three publications on the CRC screening program:

- **Thomsen, Mette K**; Rasmussen, Morten; Njor, Sisse H; Mikkelsen, Ellen M. *Demographic and comorbidity predictors of adherence to diagnostic colonoscopy in the Danish Colorectal Cancer Screening Program: a nationwide cross-sectional study*. Clinical Epidemiology 2018. Vol. 10:1733-1742.
- Mikkelsen, Ellen M; **Thomsen, Mette K**; Tybjerg, Julie; Friis-Hansen, Lennart et al. *Colonoscopy-related complications in a nationwide immunochemical fecal occult blood test-based colorectal cancer screening program*. Clinical epidemiology 2018. Vol. 10:1649-1655.
- **Thomsen, Mette K**; Njor, Sisse H; Rasmussen, Morten; Linnemann, Dorte et al. *Validity of data in the Danish Colorectal Cancer Screening Database*. Clinical epidemiology 2017. Vol. 9:105-111.

Thus, this thesis builds on my earlier work with the Danish colorectal cancer screening program, and specifically on my master's thesis. As per the specifications for the qualifying exam (i.e. the master's thesis) in the Integrated PhD and the Master's Program in Public Health Studies at Health, Aarhus University, Denmark, this thesis includes text which in earlier versions was also part of my master's thesis.

Correspondence

Mette Kielsholm Thomsen, mkthomsen.clin.au.dk / mette.k.thomsen@outlook.com

Acknowledgements

I started my path into research on my second year as a public health student, when I became a student assistant in the statistics group at the Department of Clinical Epidemiology. This was in 2014, the same year as colorectal cancer screening was initiated in Denmark. When I asked to do my BA project at the department, the screening program was an obvious area for me to engage in as a public health student. I owe a special thanks to Frank Mehnert, daily manager of biostatistics, for helping me in the right direction. I am grateful for all the possibilities offered to me at the department, and for the scientific community I have been fortunate to be a part of throughout my time here.

I was matched with Ellen M. Mikkelsen, who at the time was the epidemiologist responsible for all the clinical quality databases on cancer screening. We have had an amicable collaboration from the very first day, but Ellen has also challenged me in a way that has shaped my critical thinking and ability to communicate scientifically. Thank you, Ellen, for not only supervising me, but also mentoring me. As my bachelor's, master's and PhD main supervisor, you have urged me to assert myself, taught me about the diplomacy of research institutions and supported me in every situation throughout.

I wish to thank each of my co-supervisors for their expert advice during my PhD; specifically Rune, for moral support, for showing me a day in the clinic and for including me in the GI-team; Lars, for always being available for a quick discussion of methods and for saying "*Jamen, det gør vi da bare*"; and Henrik, for your detailed scrutiny of manuscripts. I also wish to thank my collaborators Marie D. Jørgensen and Sia K. Nicolaisen for being so pleasant to work with, and Timothy L. Lash for hosting me during my research stay in Atlanta. Finally, I wish to acknowledge the department's financial support of the project.

I have been fortunate to work alongside a long list of brilliant colleagues. Most importantly, I have always ended up with the best of office mates. We have shared ups and downs, tips and tricks, and challenges in both projects and personal life – I will settle for none mentioned and none forgotten, as we have had many different constellations in our office by the downstairs kitchen. Special thanks also to my PhD colleagues Frederikke, Anne and Sia for helpful feedback on this thesis.

Last but not at all least, Dan. I literally could not have finished this without your support and your determination to make all the ends meet in our family. Also, our two sons, Johan and Magnus, are the best, and I am grateful for how patient they have been with the thesis-writing version of me.

Thank you,

Mette K. Thomsen

Collaborators

Marie Dahl Jørgensen, MPH

Former intern and thesis student, Department of Clinical Epidemiology and Department of Public Health, Aarhus University, Denmark

Sia Nicolaisen, MSc, PhD student

Biostatistician, Department of Clinical Epidemiology, Aarhus University and Aarhus University Hospital, Denmark

Timothy Lee Lash, MPH, PhD

Professor and chair, Department of Epidemiology, Rollins School of Public Health, Emory University, Georgia, USA
Professor, Department of Clinical Epidemiology, Aarhus University and Aarhus University Hospital, Denmark

Grants

- Danish Cancer Society
- Helsefonden
- Else og Mogens Wedell-Wedellsborgs Fond
- Inge og Jørgen Larsens Mindelegat
- Kong Christian den tiendes Fond
- William Demant fonden
- Graduate School of Health

Thesis papers

The four studies included in this thesis are referred to as Study I-IV throughout, and the Roman numerals refer to the corresponding papers, Papers I-IV.

- Paper I: Mental illness and participation in colorectal cancer screening: a scoping review. Jørgensen, Marie D; Mikkelsen, Ellen M; Erichsen, Rune; **Thomsen, Mette K.** Published in Scandinavian Journal of Gastroenterology, May 2022.

- Paper II: Mental disorders and colorectal cancer screening participation and trajectories – a Danish cohort study.
  Thomsen, Mette K; Jørgensen, Marie D; Pedersen, Lars; Erichsen, Rune; Sørensen, Henrik T; Mikkelsen, Ellen M. In draft.

- Paper III: Risk-stratified selection to colonoscopy in FIT colorectal cancer screening: development and temporal validation of a prediction model.
  Thomsen, Mette K; Pedersen, Lars; Erichsen, Rune; Lash, Timothy L; Sørensen, Henrik T; Mikkelsen, Ellen M. Published in British Journal of Cancer, January 2022.

- Paper IV: Effect of a FIT-based colorectal cancer screening program on mortality estimated by the Regression Discontinuity Design.
  Thomsen, Mette K; Nicolaisen, Sia K; Pedersen, Lars; Lash, Timothy L; Erichsen, Rune; Sørensen, Henrik T; Mikkelsen, Ellen M. Submitted to American Journal of Epidemiology June 2022, resubmission ongoing.

Thesis structure

The thesis contains the following sections: a short introduction, a description of the theoretical background of screening and colorectal cancer epidemiology, a summary of existing literature and knowledge gaps, a presentation of aims of the four studies, an account of the materials and methods applied, a summary of results, a discussion of methods and results, and conclusion and implications. Lastly, the four thesis papers and other appendices are provided.

Dansk resume (Danish summary)

Screeningsprogrammer for tarmkræft er i stigende grad blevet implementeret verden over. I marts 2014 blev et screeningsprogram for tarmkræft med testen *fecal immunochemical test* (FIT) indført i Danmark for alle i alderen 50 til 74 år. Personer med en positiv FIT (over grænseværdien for hæmoglobin i afføringen) bliver tilbudt en efterfølgende koloskopi (kikkertundersøgelse af tarmen). Screening indebærer undersøgelse af en umiddelbart rask befolkning, og fordele og ulemper skal overvejes omhyggeligt. Denne afhandling indeholder fire studier som har til formål at undersøge forskellige aspekter af det danske screeningsprogram for tarmkræft, for at bidrage til den internationale evidens relevant for beslutningen om at screene for tarmkræft.

Studie I

Det er veldokumenteret at der er social ulighed i deltagelse i screeningsprogrammer for tarmkræft; men deltagelsen for personer med psykiske sygdomme er et overset område. I studie I kortlagde vi derfor den tilgængelige litteratur på området, og fandt at personer med svær psykisk sygdom deltog mindre i screening for tarmkræft. Resultaterne fra studier af mere almindelige psykiske sygdomme var tvetydige.

Studie II

I studie II sammenlignede vi deltagelsen i det danske screeningprogram blandt personer med mild/moderat og svær psykisk sygdom med personer uden psykisk sygdom. Vi fandt markant lavere deltagelse for personer med psykisk sygdom, også efter at have kontrolleret for forskelle i socioøkonomiske forhold.

Studie III

Validiteten af FIT er fundamental for screeningsprogrammet og testen skal kunne identificere personer med høj risiko for tarmkræft. Bedre risikovurdering i FIT screening er blevet efterlyst af eksperter. I studie III beregnede vi den personlige risiko for tarmkræft ud fra alder, køn og hæmoglobin-værdi. Ved at bruge en grænseværdi for den beregnede risiko fremfor FIT, ville man kunne finde en anelse flere med tarmkræft og samtidig have færre unødvendige koloskopier.

Studie IV

FIT er en forholdsvist ny metode og direkte evidens for dens effekt er begrænset. Målevariation gør det tilfældigt om en hæmoglobinværdi tæt på grænseværdien kommer over eller under. I studie IV fandt vi et fald i dødeligheden lige over grænseværdien for koloskopi.

Vi kan konkludere at det danske screeningsprogram for tarmkræft har potentiale for bedre adgang for visse befolkningsgrupper samt for bedre risikovurdering af hvem der skal tilbydes koloskopi. Vores evaluering af dødeligheden viste en beskyttende effekt af screening, men længere opfølgningstid er nødvendig for mere præcise resultater.

Summary

Screening programs for colorectal cancer (CRC) have increasingly been implemented all over the world. In March 2014, a CRC screening program using the fecal immunochemical test (FIT) was implemented for all Danish residents aged 50 to 74 years. Individuals with a positive FIT (above a cut-off value for fecal hemoglobin) are offered a subsequent colonoscopy. Screening is the examination of an apparently healthy population to detect latent disease. Careful consideration should therefore be given to benefits and harms of implementing and continuing screening programs. This thesis includes four studies aiming to investigate three different aspects of the Danish CRC screening program to contribute to the international body of evidence relevant for the decision to screen for CRC.

Study I

Social inequality in CRC screening participation is well documented; however, participation among persons with mental disorders has received less attention. In Study I, we therefore mapped the available evidence. We found that persons with severe mental disorders participated less in CRC screening, but results on more common mental disorders were ambiguous.

Study II

A lack of evidence from organized screening programs using FIT identified in Study I led us to investigate this matter in Study II. We compared participation among persons with no, mild/moderate and severe mental disorders, and found markedly lower participation among the two latter, also after having controlled for socioeconomic conditions.

Study III

The validity of the FIT is fundamental to the screening program, and the test should identify persons at high risk of CRC. Calls have been made for improved risk stratification in FIT screening. In Study III, we predicted the individual risk of CRC from age, gender and fecal hemoglobin value. A cut-off of predicted risk instead of FIT, identified slightly more cancers and caused fewer unnecessary colonoscopies.

Study IV

The FIT is fairly novel, and direct evidence of its beneficial effect is limited. In Study IV, we assumed that measurement variation makes it random if a specific fecal hemoglobin value close to the cut-off falls above or below the cut-off for colonoscopy. We identified a discontinuity (a reduction) in mortality at the cut-off as being referred to colonoscopy lowered mortality.

In conclusion, the Danish CRC screening program holds potential for greater accessibility for persons with mental disorders and for risk stratification to secure that persons referred to colonoscopy are those with the highest risk of CRC. Our evaluation of mortality showed a beneficial effect of the screening program, but a longer follow-up would provide more precise results.

Abbreviations

AN	Advanced neoplasia
ATC	Anatomical Therapeutic Chemical
AUC	Area under the curve
BMI	Body Mass Index
CI	Confidence interval
CPR	Central personal registration
CRC	Colorectal cancer
DAG	Directed acyclic graph
DCCG	Colorectal Cancer Group Database
DNPR	Danish National Patient Registry
ELSA	English Longitudinal Study of Ageing
FIT	Faecal immunochemical test
fHb	Faecal haemoglobin
FS	Flexible sigmoidoscopy
gFOBT	Guaiac faecal ocult blood test
HR	Hazard ratio
ICD-10	International Classification of Diseases, 10th edition
JBI	Joanna Briggs Institute
NPV	Negative predictive value
OR	Odds ratio
PD	Participation difference
PPV	Positive predictive value
PR	Participation ratio
PTSD	Post-traumatic stress disorder
RCT	Randomised controlled trial
RDD	Regression discontinuity design
ROC	Receiver operator characteristic
RR	Relative risk
YHS	Yorkshire health Study

Contents

Introduction 3
Background 5
Screening for cancer 5
Colorectal cancer 9
Existing literature and evidence gaps 13
Participation in screening 13
Risk-stratification in screening with fecal immunochemical test 14
Effect of screening for colorectal cancer on mortality 15
Study aims 17
Materials and methods 19
Setting 20
Data sources 20
Study designs 23
Statistical analyses 26
Results 31
Review on mental disorders and participation 31
Mental disorders and colorectal cancer screening participation 35
Predicting colorectal cancer for risk-stratified screening 37
Effect of FIT screening on mortality 39
Discussion 43
Main findings 43
Results in the context of other studies 44
Methodological considerations 46
Main conclusions 53
Perspectives and implications 55
Equal access to screening participation 55
Risk-stratified screening to prioritize colonoscopy 56
Continued evaluation and monitoring of the effect of screening 57
References 59
Appendix Table 1. Study characteristics and results of studies identified in literature search for Study III 69

Introduction

Screening for asymptomatic disease has received considerable attention within the medical field, and diseases are continuously proposed and discussed as candidates for screening. Intuitively, early detection improves treatment options and thus leads to a more favorable prognosis (1). However, this is not always the case, and as the target population for screening programs is "healthy" individuals, careful consideration should be given to the benefits and harms of implementing and continuing screening programs (2). A set of guiding principles for screening programs was first proposed by Wilson and Jungner on behalf of the World Health Organization in 1968 (3). The principles overall include knowledge about the epidemiology and natural history of the disease, suitable screening tests and infrastructure for the screening program, strong evidence for a beneficial effect of screening and continued quality control.

Colorectal cancer (CRC) is considered a disease well-suited for screening, due to a slow development and because both CRC and pre-cursers can be identified (4). Screening programs for CRC have increasingly been implemented all over the world (5, 6). In March 2014, a CRC screening program using the fecal immunochemical test (FIT) was implemented for all Danish residents aged 50 to 74 years. For individuals with a positive FIT (fecal hemoglobin (fHb) ≥20 µg/g feces), a subsequent colonoscopy is offered (7). Regardless of the type of primary screening test, the test needs to be acceptable and accessible for the target population to be both willing and able to participate. Socioeconomic inequality in CRC screening participation has been extensively evaluated and detected across countries and healthcare systems (8). However, little is known about participation of persons with mental disorders in national screening programs (9).

The validity of the screening test is fundamental to the effects of a screening program. FIT is a fairly novel screening test, and it is therefore important to investigate various aspects of its performance. Only 5.9% of participants with positive FITs in the Danish screening program turn out to have CRC, and 33% are identified with medium-risk or high-risk adenomas (10). These are important findings, however, many participants unnecessarily go through a colonoscopy procedure, which is unpleasant, expensive and entails a risk of complications (11). Recently, several calls have been made for studies to develop risk stratification methods to determine which FIT screening participants should undergo colonoscopy (12–15).

The guaiac fecal occult blood test (gFOBT), a type of stool test no longer used, has been evaluated in randomized controlled trials (RCTs) in Denmark, Sweden, the United Kingdom and the USA (16). A meta-analysis of these four trials found that screening using gFOBT decreased CRC mortality by 18%

(relative risk (RR)=0.82, 95% confidence interval (CI): 0.73;0.92) (16). On the other hand, FIT has been examined in only one cluster-randomized Chinese RCT, where FIT was combined with a risk score to select persons to undergo flexible sigmoidoscopy. This combined method showed a reduction in CRC mortality of 12% (RR=0.88, 95% CI: 0.72-1.07) compared with no screening (16). Thus, experimental data on the effect of FIT screening are very limited. Evidence on the benefit of FIT screening therefore rests on the mentioned evaluations of gFOBT and studies showing that FIT has higher sensitivity and yields higher participation than gFOBT (17–19). When screening programs are in place, observational studies based on contemporary data are needed to provide evidence on the effect of FIT screening.

The four studies included in this thesis aim to evaluate the three aspects of screening introduced above: inequality in participation according to mental disorders (Study I and II), risk stratification for selecting whom to refer to colonoscopy (Study III) and an evaluation of the effect of FIT screening on mortality (Study IV).

Background

This thesis concerns screening for colorectal cancer, and both *screening* and *colorectal cancer* come with specific conditions of importance for the research questions.

Screening for cancer

Principles of screening

Screening is defined as the examination of an apparently healthy population to detect latent disease (2). The predominant understanding of the natural history of most cancers includes a linear development from disease onset through a pre-clinical phase until onset of symptoms. Symptoms prompt the clinical phase where disease progresses or is successfully treated. Cancer screening aims to detect the disease in the pre-clinical phase and is meaningfull only when early detection leads to improved prognosis (20). However, this is not the only determinant for the appropriateness of screening. Wilson and Jungner first proposed a set of screening principles in 1968, and these principles have been widely applied and discussed since (1). Efforts have been made to consolidate the different alterations and additions to the principles through a systematic review and census process. This led to a list of principles illustrated in Figure 1.

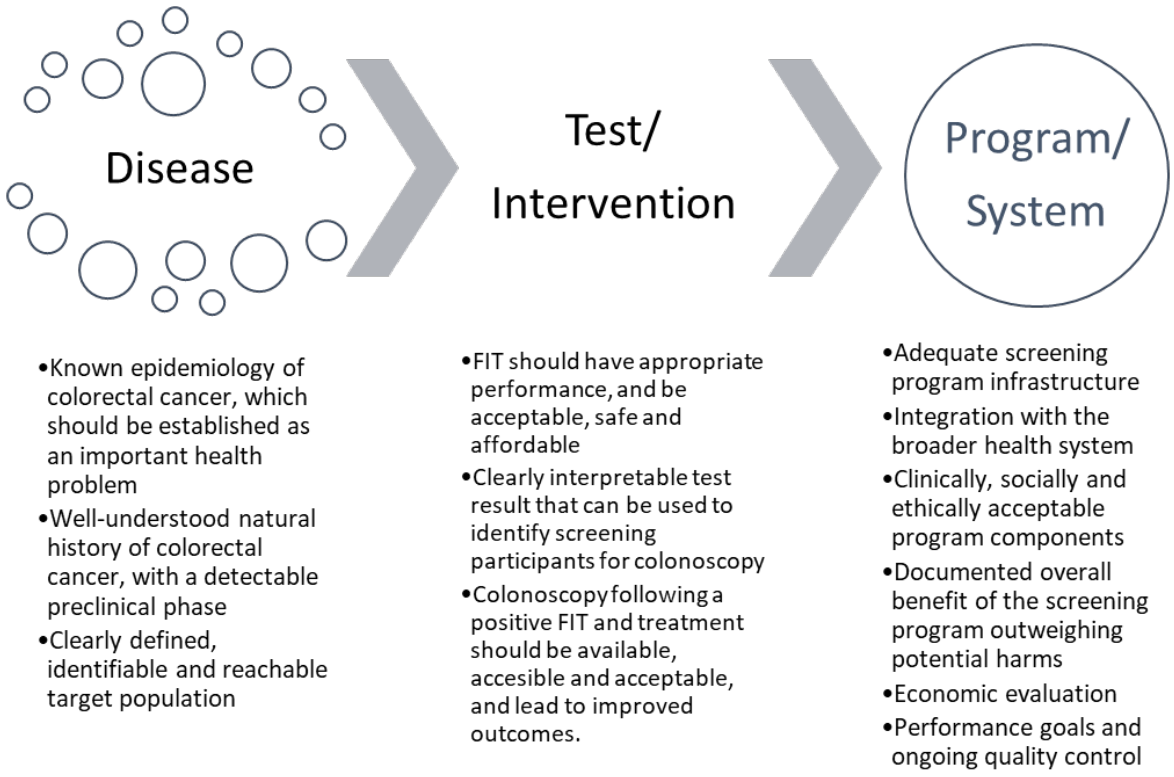

Figure 1 Principles of screening, freely condensed and adapted to colorectal cancer screening, from MJ Dobrow et al., Canadian Medical Association Journal 2018 (3)

Test validity

The set of principles for screening include the availability of a *clearly interpretable test to identify screening participants for further examinations*. However, a screening test is not a diagnostic test, and therefore it will not identify all persons with the disease; nor will it provide a negative test result to all persons without the disease. As the FIT used in the Danish CRC screening program measures fecal hemoglobin quantitatively, the choice of cut-off is essentially a matter of priority. It is necessary to balance between finding as many CRCs as possible, while keeping the amount of colonoscopy referrals at a manageable level, minimizing the number of unnecessary colonoscopies and needless stress caused for those with a false positive test result. Thus, the choice of a FIT cut-off will place the screening program somewhere between the extremes of referring everyone to colonoscopy and referring no one to colonoscopy.

The validity of a screening test can be compared to a gold standard (i.e. not a perfect test, but the best available), which can be colonoscopy in the context of FIT screening. The validity is comprised of four separate measures: sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). These calculated from the following table:

	CRC +	*CRC -*
Test +	True positive (tp)	False positive (fp)
Test -	False negative (fn)	True negative (tn)

Sensitivity is the test's ability to correctly identify those who have the disease: tp/(tp+fn). Specificity is its ability to correctly identify those who do not have the disease: tn/(tn+fp). PPV is the probability that a positive test is correct: tp/(tp+fp); and NPV is the probability that a negative test is correct: tn/(tn+fn) (2).

A test can have high sensitivity, i.e. identify many with the disease (true positives), yet have low PPV, thus cause many without the disease (false positives) to be referred to colonoscopy (21). Conversely, some false negatives will present as interval cancers due to symptoms arising before the next screening invitation (2).

Screening-specific biases

Screening changes the clinical reality as it moves time of diagnosis and introduces a cross-sectional examination of persons irrespective of whether they have symptoms or not. These conditions should be taken into account in study designs as they can create lead-time bias and length-time bias.

Lead-time bias

The aim of screening is to detect disease earlier and thereby initiate early treatment and improve prognosis. If early treatment has no life-prolonging effect, screening and early detection will only add length to the time lived with a CRC diagnosis. This corresponds to the lead-time, i.e. the time from which early detection is possible until the disease would have become symptomatic. If survival after diagnosis is compared between persons with screen-detected CRC and symptom-detected CRC, lead-time is added to the observed survival time of persons with screen-detected CRC (22).

Figure 2 depicts three different situations; A, B and C. The t_0's signifies start of follow-up, and the t_1's signifies time of death. In situation A, the person was not screened and received surgery at the usual time point, t_{0A}. This person died at t_{1A}, before the end of the fixed follow-up period (e.g. five years). In situation B, the person was screened, leading to early diagnosis and therefore earlier time of surgery, t_{0B}, than in situation A. Due to the earlier start of follow-up, i.e. t_{0B}, death at t_{1B} occurs after end of follow-up. Noticeably, however, t_{1A} and t_{1B} are identical and death was not delayed. In situation B, screening did not alter the time of death, only the time period lived with the diagnosis.

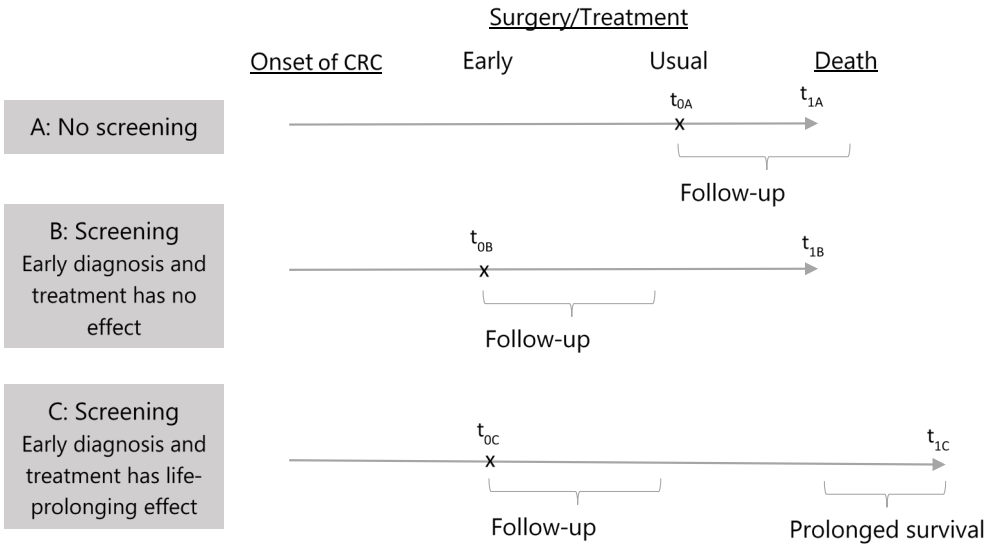

Figure 2 Survival time after diagnosis with and without screening, with or without effective early surgery – Lead-time bias. With inspiration from Fletcher, Fletcher & Fletcher 2014 (22).

In situation C, screening also led to early diagnosis and thereby early surgery. However, early surgery was effective and prolonged life, i.e. by moving t_{1C} compared with situation A and B. Like in situation B, death does not occur within follow-up. The challenge is that there is no way of distinguishing between situation B or C in a direct comparison of screening status among individuals undergoing CRC surgery. The alteration of t_0 creates lead-time bias if persons are followed since time of diagnosis or surgery/treatment.

The existence of lead-time explains why survival (e.g. 5-year survival) should not be used to estimate the effect of screening, and why survival before and after the introduction of screening cannot be directly compared. Survival is measured after a fixed interval (corresponding to the brackets in Figure 2) since time of diagnosis (t_0 in Figure 2), which was moved by screening.

Length-time bias

Length-time bias is another screening-specific bias. Screening acts as a cross-sectional examination of participants as it is conducted independently of the time of disease onset, symptoms, etc. Length-time bias is depicted in Figure 3. The arrows depict persons alive and arrowheads depict the time of death. The dashed vertical line represents the time of screening.

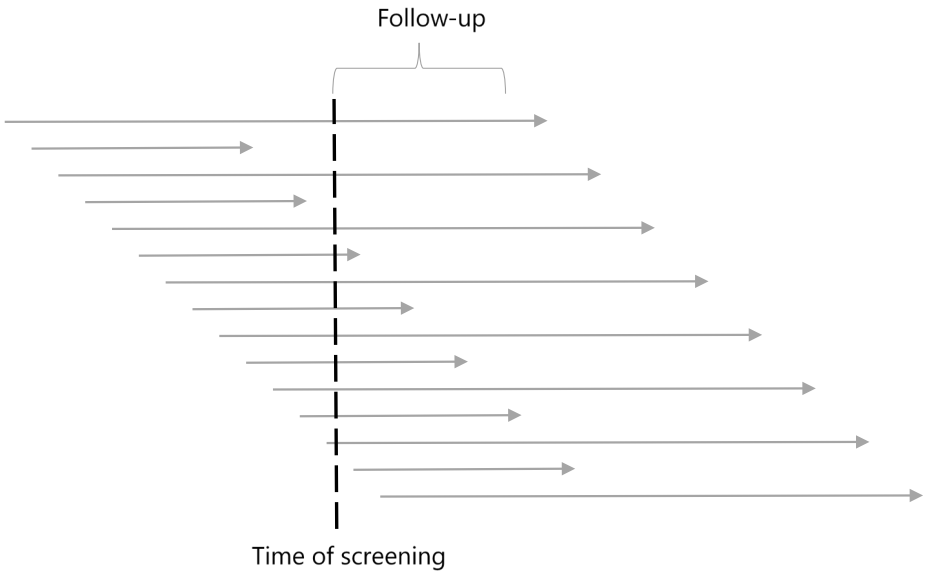

Figure 3 Cross-sectional screening – Length-time bias. At any time-point there will be more slow-progressing cancers (long arrows) identified by screening than rapidly progressing aggressive cancers (short arrows). With inspiration from Fletcher, Fletcher & Fletcher 2014 (22).

Cancers with a lengthy development (long arrows) will more likely be detected by screening as the disease more often becomes symptomatic requiring treatment or cause death before the next screening round. In contrast, rapid and often more aggressive cancers (short arrows) are less likely to be detected by screening. These basic differences in duration of the preclinical phase, and thus in the likelihood of detection by screening, justifies cautiousness with comparing screen-detected and non-screen-detected cancers. This difference also creates length-time bias. This means that persons with slowly developing cancers, which are likely to be detected by screening, may live longer after diagnosis than will persons with more aggressive cancers that are unlikely to be detected by screening; irrespective of the effect of screening and early detection.

Colorectal cancer

Colorectal cancer epidemiology

Worldwide, CRC is the third and second most common cancer in men and women, respectively. The incidence of CRC varies internationally, and the disease predominantly affects populations in western countries (23). Age-standardized incidence rates thus span from 44.8 for men and 32.2 for women per 100,000 in Australia/New Zealand to 4.5 and 3.8 per 100,000 in Western Africa (24). In Denmark, the incidence rates were 42.7 and 33.2 per 100,000 for men and women, respectively, in 2010-2014; thus Danish incidence rates are among the highest in the world (25).

The two most frequent pathways through which colorectal cancer develops are the adenoma-carcinoma pathway (70-90%) and the serrated neoplasia pathway (10-20%). Common for both is that they are initiated by genetic mutations, and precursors (adenomas or serrated neoplasia) develop into CRC. The aim of CRC screening is therefore not only to identify and treat CRC but also to identify precursors. Depending on their type and size, adenomas and serrated neoplasia are removed or monitored, thereby preventing CRC from occurring (4). Severe adenomas and CRC are often combined in the terms "colorectal neoplasia" or "advanced neoplasia" (26).

The risk of CRC increases with age, male sex and some colon diseases, such as familial tendency for polyps, Crohn's disease, and ulcerative colitis (23). Thus in 2013, the incidence of CRC in Denmark per 100,000 ranged from very low below age 40 to 103.6 for men and 80.9 for women at age 55-59 to 604.6 for men and 396.6 for women at age 80-84 (27). Proximal risk factors include smoking, alcohol consumption, diet (high in red/processed meat, low in fruit, vegetables and fibers) and physical inactivity. CRC typically develops over many years, and small amounts of blood in the stool is an early sign of (pre)cancerous lesions (2). Symptoms are weak (e.g. change in bowel habits and unexpected weight loss) and present at a late stage. Survival depends on stage at diagnosis (28, 29). In Denmark, the age-standardized five-year survival has been steadily increasing from 44% for men and 48% for women in 1992-1996 to 60% for men and 62% for women in 2007-2011 (30).

Implications for lead-time and length-time

The progression to colorectal cancer happens slowly; over an estimated period of 10 to 15 years (4). This means that length-time bias may work counterintuitively for CRC screening. Over-diagnosis caused by screening can be seen as an effect of length time in the extreme, i.e. the diagnosis of very slow cancers that would never develop enough to contribute to death. Thus, mortality among screening-detected cancers could be underestimated due to over-diagnosed cancers, which would

never have been detected in the absence of screening. The long development also means that lead-time could cause considerable bias, as the time from which CRC can be detected by screening tests until symptoms arise can span several years (31).

Screening tests for colorectal cancer

Overall, two screening modalities are in use for early detection of CRC: endoscopy (colonoscopy or flexible sigmoidoscopy) and stool tests detecting signs of blood in the stool, followed by endoscopy (5, 6). Stool sample tests include gFOBT and FIT. The gFOBT is a qualitative test that requires several samples of stool from consecutive bowel movements. RCTs of gFOBT started in the 1970s. Since then, organized and opportunistic CRC screening have been widely implemented; Italy, Israel and Japan initiated screening programs in the 1980s and 1990s; and many other countries have followed, especially in the end of the 2000s (5).

Testing techniques have since improved, and the FIT was developed in the late 1990s and early 2000s (18, 32). The FIT can be qualitative or quantitative in the measurement of fecal hemoglobin. It only requires a single stool sample, and therefore has higher participation and sensitivity in population-based programs than the gFOBT. This has led many screening programs to use this test instead of the gFOBT. CRC screening modalities under development include blood sample tests (33) and colon capsule cameras (34).

Colonoscopy after a positive FIT

A colonoscopy is an endoscopic examination of the bowel, and requires full laxative bowel preparation to empty the colon (35). The quality of the colonoscopies performed after screening is important for effect achieved of the screening program. The ultimate measure of colonoscopy quality is the rate of interval cancers, i.e. cancers missed by screening, but detected due to symptoms before the next screening. However, process performance measures include the adenoma-detection rate of each colonoscopist (36). Thus, the quality of a colonoscopy is both determined by the experience and capabilities of the colonoscopist, and also of adequate bowel preparation by the patient.

The colonoscopy procedure often entails discomfort to the patient, and it has a small risk of serious complications, such as perforation of the colon, bleeding or medical complications (11). Out of all colonoscopies after a positive FIT in the Danish screening program in 2020, serious complications occurred in 0.21% (95% CI: 0.15;0.28) (37).

The colorectal cancer screening program in Denmark

The Danish population-based CRC screening program for all residents aged 50 to 74 years was initiated in 2014 with a four-year implementation round. Subsequently, screening has been biennial; and currently the fourth screening round is on-going (38). The CRC screening program is part of the tax-funded Danish healthcare system, and screening is offered free of charge at the point of use (39). The program uses FIT followed by colonoscopy referral for those screening positive. The first round of screening is also called the prevalence round as the CRCs identified may have been developing for several years, whereas in the subsequent screening rounds identified cancer will increasingly be incident (2).

Figure 4 shows an overview of the Danish CRC screening program with measures of participation, positive FITs, colonoscopy adherence and identified CRC reported by the Danish CRC screening database (10, 37).

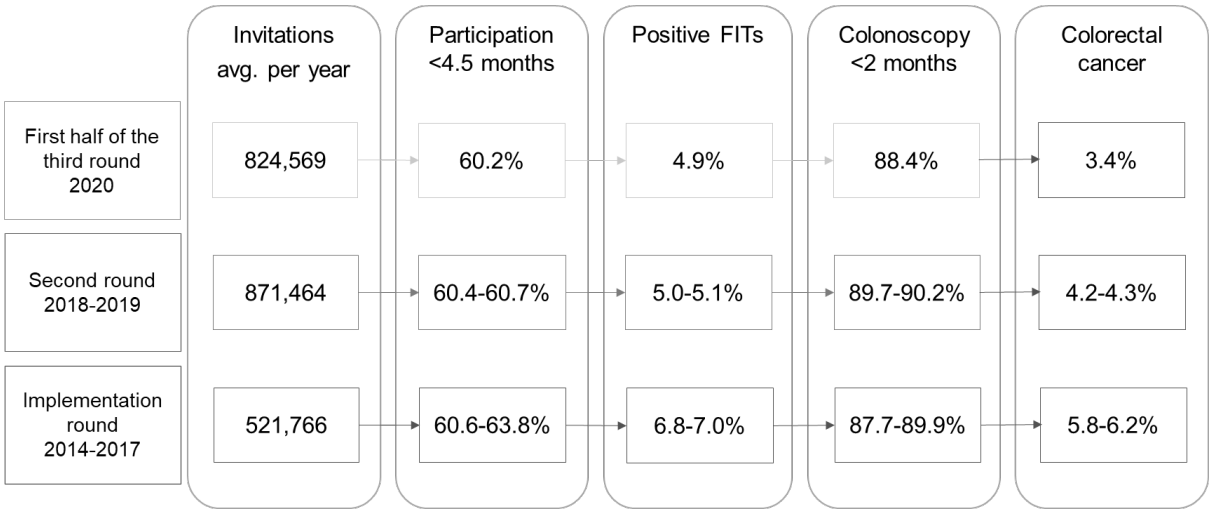

Figure 4 Invitations, participation, positive FITs, colonoscopy adherence and findings of CRC in the first, second and first half of the third round of the Danish CRC screening program. Danish CRC screening database (10, 37).

Further details of the invitation procedure, screening test, etc., are described under the subheading "Setting" in the methods section.

Existing literature and evidence gaps

The population-based Danish CRC screening program is of comparatively recent date as has, it is an emerging research field in Denmark. Several research groups have evaluated various aspects of the national screening program since its initiation in 2014, and the body of evidence on the program is steadily growing. So far, 18 papers have been published broadly related to the three aspects of CRC screening studied in this thesis; participation, risk-stratification and effect on mortality:

- Screening participation by gender, age, socioeconomic position, marital status (and partner concordance), immigration status, health literacy and health behavior (40–46)
- The effect of reminders on participation by gender, age, income, educational level, marital status, and immigration status (47)
- Risk of ineligible (non-analyzable) stool samples by immigrations status (48)
- Sensitivity and specificity at different levels of FIT cut-off (49)
- Adherence to the subsequent diagnostic colonoscopy by gender, age, socioeconomic position, marital status, immigration status and somatic morbidity (50, 51)
- Barriers towards undergoing colonoscopy after a positive FIT, examined in a qualitative interview study (52)
- CRC and adenoma findings (53), including cancer stage distribution (54), among persons invited vs. not-yet-invited to CRC screening
- CRC mortality after screening among persons invited vs. not-yet-invited (55) and pre- vs. post-screening initiation (56) and within 90 days of surgery comparing screen-detected and symptom-detected CRCs (57)

No studies have evaluated the role of mental disorders in screening participation or addressed the potential of using risk stratification in selection for colonoscopy in the Danish CRC screening program. Three studies have evaluated the effect of screening on mortality using three different study designs. In the following, I present the existing evidence in relation to Study I-IV.

Study I+II

Participation in screening

Participating in CRC screening is not only a choice determined by preference for screening or no screening. It is also affected by the type of healthcare system (e.g. insurance requirements), organization of the screening program and individual resources (2). A 2016 review including published and unpublished information from organized CRC screening programs

worldwide consistently found that disadvantaged groups with the lowest education, lowest income or highest deprivation score, etc., had reduced participation in CRC screening across programs (8).

However, conditions other than socioeconomic disadvantage may affect the ability to participate in CRC screening. I hypothesized that mental disorders could also affect participation. Acutely being affected by severe mental disorders such as psychosis or mania may can reduce the capacity for decision making, and this could include the decision of whether to participate in screening (58). Regardless of severity, persons with mental disorders may experience more barriers to screening (59).

I found only one existing review, by Solmi et al., on mental disorders and participation in all types of cancer screening. Solmi et al. found no association between mental disorders and participation in their meta-analysis of CRC screening (60). However, they used narrow inclusion criteria for their review and included only eight studies on CRC screening. They did not go into depths with findings for specific cancer types, and therefore we conducted a scoping review (Study I), with broader inclusion criteria, aiming to map all types of existing literature and identify specific gaps in evidence on participation in CRC screening among persons with mental disorders (9). In their meta-analysis, Solmi et al. included one study from Europe, six from the USA and one from Australia with "any mental disorder" as the exposure. They found no studies on participation among persons with schizophrenia, and only the study from Europe and four of the studies from the USA reported specifically on participation among persons with mood disorders. Thus, the limited evidence led us to conduct Study II, investigating participation among persons with and without mental disorders in the Danish CRC screening program.

Study III

Risk-stratification in screening with fecal immunochemical test

CRC screening has challenged colonoscopy capacity in many countries, including Denmark (12, 14, 51, 61, 62). To reduce the colonoscopy burden, some studies have focused on identifying an optimal cut-off value for fecal hemoglobin (49, 63) or explored stratified cut-offs for men and women (64–67). A dose-response relationship exists between fecal hemoglobin and the risk of CRC (68). Even though the FIT provides a quantitative measurement of fecal hemoglobin, a dichotomized FIT cut-off is used that implicitly assigns the same high risk of CRC to all values above the cut-off. Including information on the fecal hemoglobin value and other factors as predictors in a prediction model may allow for more detailed risk stratification. A prediction model should be evaluated according to its discrimination and calibration, preferably also in an external dataset (69). To identify relevant existing literature, I applied the following criteria:

- Models predicting CRC or advanced neoplasia
- Quantitative FIT included as a predictor
- Performance evaluation of discrimination and/or calibration

From a search string corresponding to these criteria, I identified three relevant studies; by Yen et al., Stegeman et al. and Cooper et al. (70–72). A search string and an overview are provided in Appendix Table 1. Yen et al. presented three prediction models for advanced neoplasia; one with proximal risk factors, e.g. gender, Body Mass Index (BMI) and family history of CRC; one with fecal hemoglobin as a single predictor; and one with all predictors combined. They evaluated discrimination and conducted external validation, and their model performed well in both. However, their study population included persons with both positive and negative FITs, so they combined advanced neoplasia missed by FIT screening and screen-detected advanced neoplasia into one outcome. In addition, outcomes occurred a mean of 4.25 years after screening participation, and methods were generally not clearly described (72). Stegeman et al. combined fecal hemoglobin with the following risk factors; age, gender, smoking status, family history of CRC and calcium intake. They found that their prediction model stratified better according to risk of advanced adenomas compared with using a FIT cut-off. They evaluated both discrimination and calibration but did not externally validate their model. However, this was the only study to directly quantify the reduction in the number of false positives (73). Cooper et al. presented a model including age, gender, participation in past screening and fecal hemoglobin to predict advanced neoplasia. Their model was well calibrated but had limited discriminatory power and had not been externally validated. Sensitivity was higher for their prediction model than with FIT cut-offs (71).

Overall, these studies suggest a potential for risk stratification using prediction models in FIT screening; however, this has not been explored in the context of the Danish CRC screening program.

Study IV

Effect of screening for colorectal cancer on mortality

No trial has compared FIT with no screening (74). A meta-analysis of four RCTs comparing gFOBT with no screening found a null effect on all-cause mortality (RR=1 (95% CI: 0.99;1.01)) and a 13% reduced CRC mortality (RR=0.87 (95% CI: 0.82;0.92)) after 17 to 23 years of follow-up. Since this meta-analysis was conducted, a small number of studies have evaluated the effect of FIT screening on mortality using observational designs (55, 56, 75). In a multinational European study, Cardoso et al. estimated changes in incidence, stage distribution and mortality of CRC over time. The study reported CRC mortality in Denmark in 2002-2015 and found an annual change of -2.3% (95% CI: -2.7;-1.9) for men and -2.5% (95% CI: -3.0;-2.0) for women (56).

However, this mortality reduction cannot be attributed to screening due to the short period elapsed since screening was initiated in early 2014. By the end of 2015, no more than half of the target population would have been invited to participate in screening as the first screening round lasted four years. The study observed the largest decreases in CRC mortality in countries with the earliest implementation of screening programs (opportunistic colonoscopy screening in Austria, Germany and the Czech Republic), less so in countries with later screening implementation (including Denmark) and least so in countries with no CRC screening (e.g. Norway and Sweden). A major limitation of this study is its pre-post design, which does not allow the authors to fully account for the effect of better treatment, etc., over time. In addition, the interpretation of the study is also hampered by pooling data on various screening modalities across countries and across periods since screening initiation. Another ecological study compared the development in CRC mortality between areas with early vs. late FIT screening initiation from 2002 onwards in Italy. They reported a larger change in the trend of CRC mortality after 2002 with a rate ratio of 0.97 (95% CI: 0.95;0.98) in areas with early vs. late initiation (75).

In a recent study from Denmark, Njor et al. compared CRC mortality among invited vs. not-yet-invited persons (55) utilizing the random sequence of birth months for invitations in the implementation round (2014-2017) of the screening program. The design is therefore specific to the implementation screening round, and follow-up was limited (median 3.3. years). Njor et al. found a RR of 0.83 (95% CI: 0.66;1.03) for CRC mortality, and did not estimate the effect on all-cause mortality. Stratified by sex and age, the largest effect was found for 60-71-year-old men (RR=0.68 (95% CI: 0.49;0.94)). Secondary to their intention-to-treat analysis, they also compared participants within the invited group to a group of not-yet-invited while adjusting for healthy user bias (using the RR between non-participants and those not-yet-invited as the correcting factor). This analysis estimated a RR of 0.71 (95% CI: 0.46;1.08).

Other observational studies, including a study of the effect on mortality of the Danish CRC screening program, failed to grasp the influence of screening-specific biases (57, 76–80). All of these studies estimated the effect of screen-detected CRC versus symptom-detected CRC on mortality in patients who underwent CRC surgery. As mentioned, a comparison of participants with non-participants is affected by both lead-time bias and length-time bias, and therefore likely overestimates the effect of screening. The limitations of ecological studies, the non-applicability of approaches comparing invited and not-yet-invited participants after the implementation screening round, and not least the bias in comparisons of participants and non-participants warrant new study designs to duly evaluate the effect of CRC screening on mortality.

Study aims

The overall aim of this PhD project is to address issues regarding aspects of participation, risk-stratification and effect of screening in the Danish colorectal cancer screening program, and thereby contribute to the international body of evidence relevant for the decision to screen for colorectal cancer.

The thesis includes four studies, with the following specific aims:

Study I

Aim: to summarize the literature on CRC screening participation among persons with mental disorders and identify knowledge gaps.

Study II

Aim: to evaluate the association between mental disorders and participation in colorectal cancer screening, as well as describe trajectories in the screening program after a positive FIT.

Hypothesis: persons with severe and mild/moderate mental disorders participate less in CRC screening, and the inequality will persist after adjustment for socioeconomic conditions.

Study III

Aim: to develop and validate a prediction model incorporating age, gender and fecal hemoglobin to predict individual risk of CRC in screening participants with a positive test.

Hypothesis: the prediction model will identify more cancers and adenomas and reduce the number of unnecessary colonoscopies compared with FIT alone.

Study IV

Aim: to evaluate the effect of a positive FIT on all-cause mortality and CRC mortality using the quasi-experimental regression discontinuity design.

Hypothesis: testing just above the FIT cut-off, and therefore being referred to colonoscopy, will lower CRC mortality and all-cause mortality compared with those testing just below the cut-off.

Materials and methods

In Table 1, I provide an overview of the applied study designs, data sources and analysis (excluding Study I due to its review design). Further details on variable definitions, secondary analysis, etc., are supplied in Paper II, III and IV.

Table 1 Overview of materials and methods for Study II-IV.

	Study II	Study III	Study IV
Aim	To evaluate the association between mental disorders and participation in colorectal cancer screening.	To develop a prediction model for risk-based selection of participants to undergo diagnostic colonoscopy.	To evaluate the effect of a positive FIT on all-cause and CRC mortality using the regression discontinuity design.
Study population	Persons invited from March 2014 to September 2018. n=2,036,352	Persons invited in 2014-2015 (development) and 2016 (validation) who had a positive FIT and colonoscopy. n=56,459	Persons invited in 2014-2019 with an analyzable FIT 14–<26 µg fHb/g feces at their first screening. n=35,353
Study design	Cohort study	Cross-sectional prediction study	Cohort study Regression discontinuity design
Data sources	CRC screening database, Danish National Patient Register, Prescriptions Registry, Health Service Contacts Registry, Danish Civil Register, Education Register	CRC screening database, Danish colorectal cancer group database	CRC screening database, Danish National Patient Register, Health Service Contacts Registry, Danish Civil Register, Cause of Death Register, Education Register, Labour Classification Module
Exposures/predictors	Mental disorders	Fecal hemoglobin value, gender, age	Fecal hemoglobin value
Outcomes	Participation Trajectories: opting out, FIT results, colonoscopy adherence, completeness	Colorectal cancer, medium/high-risk adenomas, low-risk adenomas	Mortality, CRC mortality
Statistical analysis	Generalized linear models using pseudo values to estimate risk differences and RRs at a fixed time point.	Risk prediction: Logistic regression Performance: AUC and calibration slope. Validation: Temporal, persons invited in 2016.	Cox regression estimating hazard functions on each side of the cut-off and the ratio between them.

Abbreviations: AUC=area under the curve, CRC=colorectal cancer, fHb=fecal hemoglobin, FIT=fecal immunochemical test, RR=relative risk.

Setting

Figure 5 Sample tube for the fecal immunochemical test

The Danish CRC screening program was initiated in 2014 and is offered biennially to all residents aged 50-74 years. The implementation round lasted four years until the end of 2017 and involved inviting the target population in a randomized order of birth months, except for individuals entering and leaving the age range for CRC screening; they had to be invited one month before their 50th or 75th birthday at the latest.

Test-kits for home-sampling and an invitation letter are mailed directly to those who are invited to screening. The administration of the screening program, including invitations, shipment of sample kits, return of screening results, and booking of colonoscopy, is automated and handled by a central system owned by the five Danish Regions.

The program uses a FIT of the brand OC sensor from Eiken, Japan, with a 20 µg fHb/g feces cut-off for a positive screening test. Persons screening positive are referred to colonoscopy within 14 days according to a national cancer treatment guarantee. Participants with a negative FIT, i.e. <20 µg fHb/g feces, are reinvited in the next screening round two years later (38). Persons with a positive FIT, but a colonoscopy with no findings (called clean colon) will have their next invitation postponed by eight years (81). This is because a clean colon colonoscopy has been associated with lower odds of CRC up to ten years after the colonoscopy (odds ratio (OR)=0.23 (95% CI: 0.19;0.27)) (82).

Data sources

Study II, III and IV were register based, utilizing data from existing administrative registries and clinical quality databases, linked by the central personal registration (CPR) number. Data sources included the Danish CRC screening database, the Colorectal Cancer Group Database (DCCG), the Cause of Death Registry, the Danish National Patient Registry (DNPR), the Prescriptions Registry, the Health Service Contacts Registry, and demographic and socioeconomic data tables from Statistics Denmark. An overview of the data sources is shown in Figure 6, along with the description of data retrieved from each source.

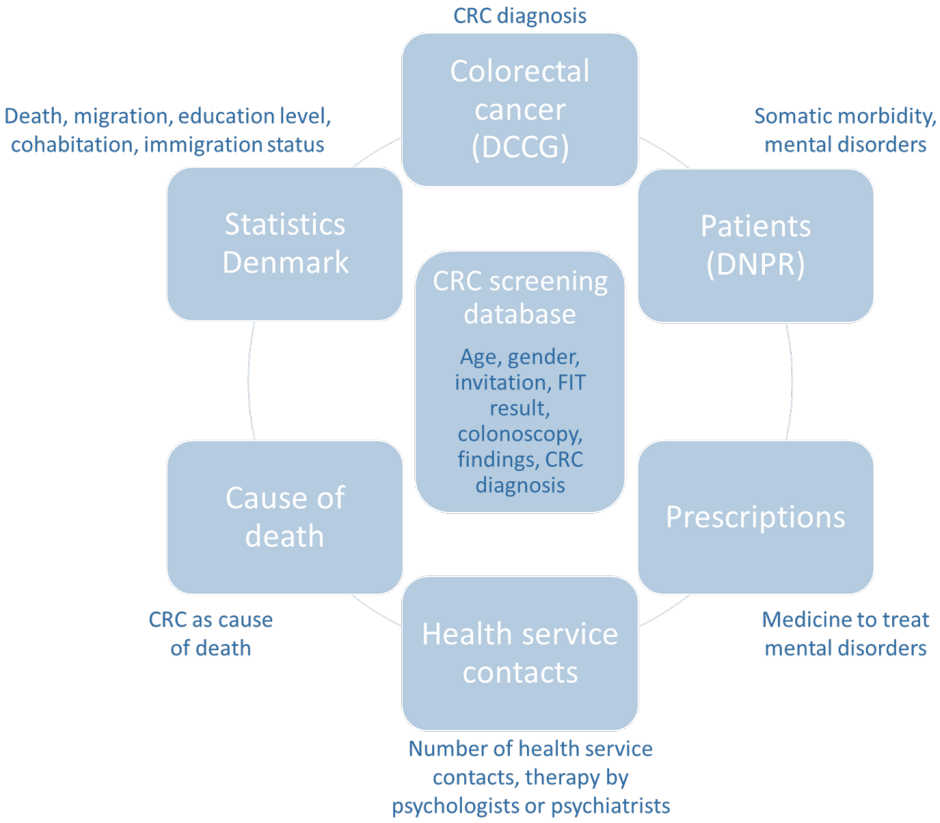

Figure 6 Included data sources and the types of information retrieved.
Abbreviations: CRC=colorectal cancer, DCCG=Danish Colorectal Cancer Group database, DNPR=Danish National Patient Registry, FIT=fecal hemoglobin test.

CRC screening database

Information on all steps of the screening program, i.e. invitation date, FIT results and colonoscopy findings, is gathered in the Danish CRC Screening Database from existing registries: the regions' invitation and administration module, the DNPR and the Pathology Registry. This clinical quality database is updated daily from the Civil Registry, and therefore the population included is complete. Together with colleagues, I validated a sample of the first DNPR data in the CRC screening database, as some novel procedure codes were implemented with the screening program. We found high validity of the existing variables, whereas some of the newly implemented screening-specific procedure codes (not used in this thesis) had lower validity (38). Study populations for Study II-IV were identified in the CRC screening database.

Colorectal cancer (DCCG)

The DCCG is a clinical database. It includes all persons with CRC treated at surgical departments in Denmark. The database has high completeness for information on patients and cancer diagnosis and treatment at the time of diagnosis (83). Few diagnoses of CRC may be missing in the DCCG, but these

diagnoses may be registered in the DNPR and vice versa; thus, I used both to identify CRC cases in Study III.

Patients (DNPR)

The DNPR contains dates, diagnoses and procedures for hospital contacts since 1977 (84). The somatic and psychiatric health systems are reported similarly, but in separate data sources for research. I used the psychiatric DNPR to identify persons with a diagnosis of mental disorders, registered at both in-patient and out-patient contacts. In addition, I used the DNPR to define somatic comorbidity from diagnosis codes registered at in-patient and out-patient contacts in the somatic health system.

Prescriptions

Health service contacts

The prescriptions registry contains all prescriptions redeemed at Danish pharmacies with the Anatomical Therapeutic Chemical (ATC) code for each medication (85). The Health Service Contacts Registry contains all contacts paid or subsidized by the state health insurance scheme (39). Both these registries were used to identify persons in treatment for mental disorders in Study II, and the overall number of contacts to health practitioners was calculated from the Health Service Contacts Registry to describe the study population in Study IV.

Cause of death

The Cause of Death Registry includes up to four conditions contributing to cause of death and up to eight conditions existing at the time of death, as registered by a physician on the death certificate. Until 2002, causes of death were coded from death certificates by specialist, trained coders under the National Board of Health. However, this practice was abandoned to reduce the time lag of data availability. Now, no central validation is performed as conditions are classified according to the International Classification of Diseases, 10th edition (ICD-10) codes and submitted digitally by the individual medical doctor issuing the death certificate (86). Cause-of-death data are considered of suboptimal quality due to uncertainties in determining a main underlying cause of death (86). To reduce this problem, I included all levels of codes in the Cause of Death Registry for the definition of CRC death in Study IV.

Statistics Denmark

Potential confounders in Study II and population characteristics in Study IV were retrieved from socioeconomic registries maintained by Statistics Denmark. Data tables included: Demography, Vital status and migration, Education, and the Labour Classification Module. These registries are all highly complete national administrative registries (39). However, the datasets are created for persons alive and residing in Denmark at a specific date, e.g. 31st of December for the Education register.

Therefore, persons who died during a specific year will not be in that year's dataset, and I therefore used data from the previous year.

Study designs

Study II-IV concern different stages and study populations in the screening program, as outlined in Figure 7. This shows how the target population is invited, and it shows various trajectories through the screening program. Participation can lead to a positive (≥20 µg fHb/g feces), negative or inconclusive FIT result. In case of a positive FIT result, the participant is referred to colonoscopy, where the colon is examined for abnormalities. If the FIT is negative, the participant will be re-invited in the next screening round.

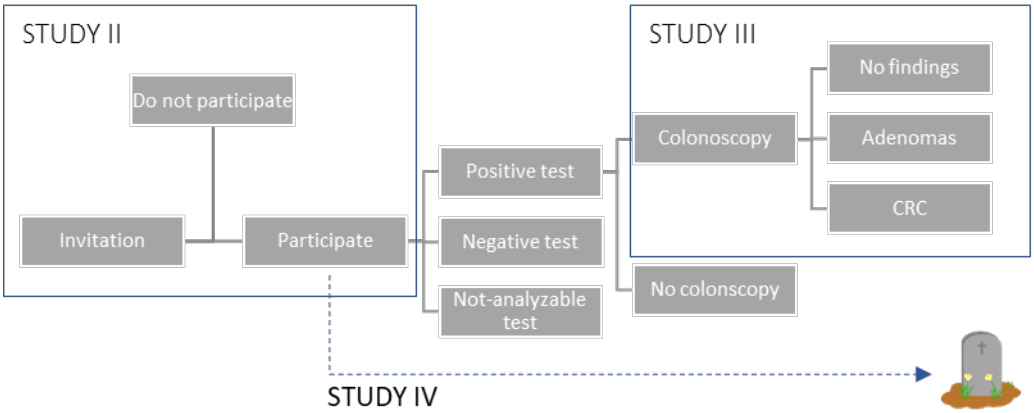

Figure 7 Study designs related to the flow of the screening program

Study II concerns participation among all persons invited to screening, Study III concerns findings among persons undergoing colonoscopy, and Study IV follows participants to estimate mortality. We updated the dataset during the project period, and therefore the study cohorts are included from different time periods. Descriptions of each study design follow below, including the methods used to conduct the review of existing literature in Study I.

Study I

Scoping review

Study I was conducted as a scoping review, which is a relatively new type of review that until recently had no universal definition or methodology (87). In 2015, Peters et al. published a guide to conduct scoping reviews, which are especially useful when a topic has not previously been subject to a comprehensive review and if the body of literature is heterogeneous and therefore not suited for systematic review (87). We aimed to map all types of available evidence and identify gaps in the existing literature. In 2018, Tricco et al. published an extension to the systematic review guidelines, developed by international experts: the Reporting Items for Systematic Reviews and

Meta-analyses: extension for scoping reviews (PRISMA ScR) (88). We adhered to these guidelines in the reporting of Study I (9).

Broad searches in different databases were conducted by one author to identify papers for review, see details in Paper I. Studies on participation in CRC screening among persons with mental disorders were included for review, and we excluded studies on persons with intellectual and developmental disabilities and dementia.

Scoping reviews often include all studies on a subject, irrespective of their quality, and therefore quality assessment is not a requirement in a scoping review (87). However, we assessed the quality of the included studies using the Joanna Briggs Institute's (JBI) critical appraisal tool (89). The tool consists of questions related to, e.g., recruitment and sampling, exposure and outcome definitions, confounding and statistical methods. The quality assessment was independently performed by two authors and any discrepancies were resolved among us.

Study II

Cohort study on mental disorders and participation

Study II is a cohort study including all residents invited to the Danish CRC screening program from March 2014 to September 2018. Some studies on CRC screening participation have not accounted for censoring or competing risk throughout follow-up (45, 90, 91). However, as it is well-documented that persons with mental disorders have a high excess mortality (92, 93), we therefore included all invited persons and used a regression model taking censoring from emigration and competing risk from death into account.

Mental disorders

We defined severe mental disorders as any hospital discharge diagnosis of schizophrenic disorders, psychosis, mania and bipolar disorders, depression, anxiety, obsessive compulsive disorder, stress-related disorders, somatoform disorders, eating disorders, and attention-deficit disorders or ≥2 prescriptions for antipsychotic medications according to codes specified in supplementary Table 1 for the paper. Mild/moderate mental disorders were identified by ≥2 psychologist/psychiatrist therapy contacts or ≥2 prescriptions of antidepressive or anxiolytic medications. Each person's main exposure was categorized according to the most severe disorder assessed in the five years preceding screening invitation.

Outcomes

Participation in CRC screening was defined as having a date for a registered FIT result in the screening database. To study trajectories after participation, we included the following outcomes: degree of opting out (permanently or temporarily), FIT results, non-analyzable tests (if there was no subsequent positive/negative FIT test), colonoscopy adherence, and completeness of colonoscopy.

Confounders

To determine which confounders to adjust for, I mapped the relation between mental disorders and screening participation in a directed acyclic graph (DAG), see Figure 8.

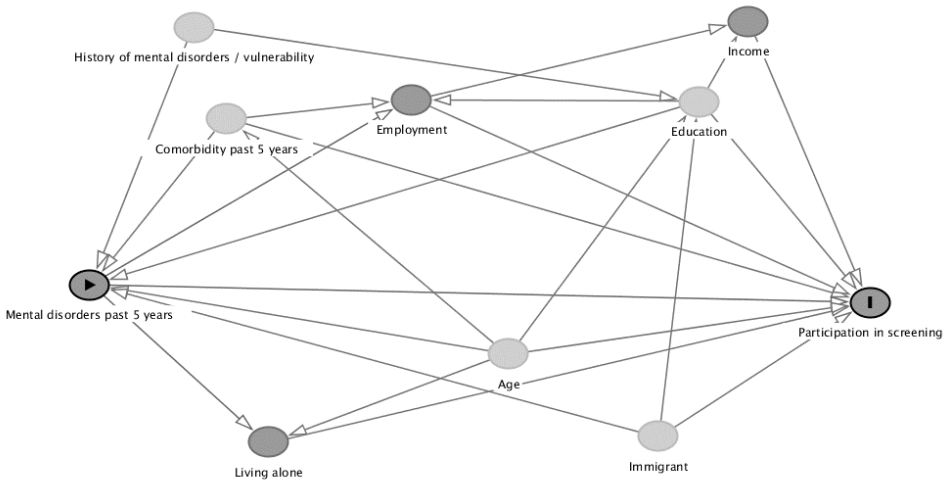

Figure 8 Directed acyclic graph of the relation between mental disorders and participation in screening. Made using daggity.net.

I considered demographic and socioeconomic conditions in addition to somatic comorbidity as potential confounders. Following Bambra's 2011 paper on social inequalities, I focused on absolute disadvantage (e.g. low income, and none or low education) instead of gradients of inequality (e.g., years of education) for socioeconomic conditions (94). Assuming that age, being an immigrant and low education level affect both the risk of mental disorders and screening participation, while unemployment, low income and living alone were descendants of low education and mental disorders, I specified an adjustment set consisting of: age, education level, immigration status and somatic comorbidity.

Study III

A prediction model for risk stratified screening

Study III included those who had a positive FIT (≥20 µg fHb/g feces) and underwent colonoscopy thereafter among all persons invited to screening in March 2014 to December

2016. We developed a model predicting the findings of CRC and adenomas to risk stratify the selection to colonoscopy.

As predictors, we included age, gender and fecal hemoglobin value; information that is all currently available in the screening program, so no further data collection was needed for the prediction model.

Conducting prediction studies is a sub-discipline within clinical epidemiology, with its own terminology and quality guidelines. The methods applied relied on the 2009 series of articles on "Prognosis and Prognostic research" in the *British Medical Journal* (95–97), the 2015 TRIPOD statement ("Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis") (69), and papers by Steyerberg & Vergouwe (98) and Steyerberg et al. (99). These methodological papers stress the need to evaluate the performance of a prediction model by both discrimination and calibration, which is conducted through comparisons of the predicted vs. the true outcomes.

Study IV

The regression discontinuity design to estimate mortality after screening

Study IV included those among persons invited to screening in March 2014 to December 2019 who participated with an analysable FIT. We restricted the main analysis to persons who had a FIT result in a narrow range around the cut-off (17–<23 µg fHb/g feces). In addition, we explored using a wider fecal hemoglobin range (14–<28 µg fHb/g feces). The end of follow-up was 2020 for all-cause mortality and 2019 for CRC mortality due to limited data availability. Accordingly, we had four study populations; one in each range for both all-cause mortality and CRC mortality.

We used the regression discontinuity design (RDD) to evaluate the effect on mortality of having a positive FIT and being referred to colonoscopy. The RDD has been widely used in economic research; but while the design is underutilized in epidemiology, it holds great potential for causal estimation (100). The RDD is a quasi-experimental design, i.e. a method that makes use of a mechanism resulting in randomized assignment of an exposure or intervention. With the RDD, the intervention is not randomly assigned by the investigators like in a RCT but by measurement variation around the cut-off of a continuous variable. Thus, the RDD can be used when a cut-off is applied to a continuous measure (here fecal hemoglobin) to determine whether to offer an intervention (here colonoscopy) or not.

Statistical analyses

A summary of the statistical analyses in each study is provided below, and further details are included in Paper II, III and IV.

Study II

The pseudo observations method for longitudinal data

I estimated absolute differences and ratios for participation (yes=1, no=0) using the pseudo observations method to handle competing risk from death (101). I first computed pseudo values for the cumulative incidence function with competing risk and then included the pseudo values as outcomes in generalized linear equations to estimate absolute differences and ratios of participation (101).

In all analyses, I compared outcomes in persons with and without mental disorders with "no mental disorders" as the reference. The cumulative incidence was adjusted for age, education level, immigration status and somatic comorbidity. All analyses were stratified by gender. To describe differences in trajectories after participation, I calculated proportions of participants with 1) positive, negative or non-analyzable FITs, 2) persons with positive FITs who underwent colonoscopy within 60 days and 3) colonoscopies that were incomplete.

Study III

Logistic regression to develop a prediction model

I included age, gender and fecal hemoglobin level in logistic regression models to estimate ORs of CRC and advanced neoplasia. I evaluated the linear fit of the continuous variables (age and fecal hemoglobin values) with restricted cubic splines. Fecal hemoglobin values were truncated above 200 µg fHb/g feces which is the maximum measured FIT value. Therefore, the fecal hemoglobin value did not have a linear fit, and I included fecal hemoglobin as a categorical variable.

I also evaluated interactions among the three variables by including an interaction term in the regression analysis. Spline and interaction terms were evaluated using a Wald test of the null hypothesis for each expansion (with a limit of 5% type 1 error).

Prediction model quality and performance

Prediction model performance was evaluated by discrimination and calibration. Discrimination describes the ability of a model to assign a higher risk to persons with disease than to disease-free persons. Model calibration assesses the accuracy of predicted probabilities compared with observed outcomes, such as whether 10% of those with 10% predicted risk had the outcome (96). Discrimination and calibration can be assessed in the original dataset for development purposes; however, the performance of prediction models should also be validated externally. This can, e.g., be done in a study cohort from another country or from a separate medical center. Such a data resource was not available for Study III, but I conducted a temporal validation, which is considered an intermediate between an internal and an external validation (97, 102). Thus, I divided data into two

datasets of participants invited in 2014–2015 (development dataset) and 2016 (temporal validation dataset).

To compare the ability of the prediction model to risk stratify participants, I calculated the predicted risk of CRC and advanced neoplasia for each person. As the study population included only those who had a positive FIT, we compared the two risk cut-offs with a FIT cut-off using a hypothetical scenario in which the number of colonoscopies was reduced corresponding to the colonoscopies initially anticipated based on 5.2% positive FITs (103). Findings of no abnormalities, low-risk adenomas, medium/high-risk adenomas and CRCs were compared between the risk cut-offs and FIT cut-off.

Study IV

Cox regression with a discontinuity

The RDD rests on the assumption that individuals with values immediately below or above the cut-off are randomly distributed to either side because of random variability in the FIT measurement (104). Both observed and unobserved variables (confounders) should also be randomly distributed and the groups should thus be exchangeable. To test this assumption for observed variables, I calculated proportions of age, gender, employment, education and health service contacts among persons with FIT results on each side of the cut-off in the two ranges.

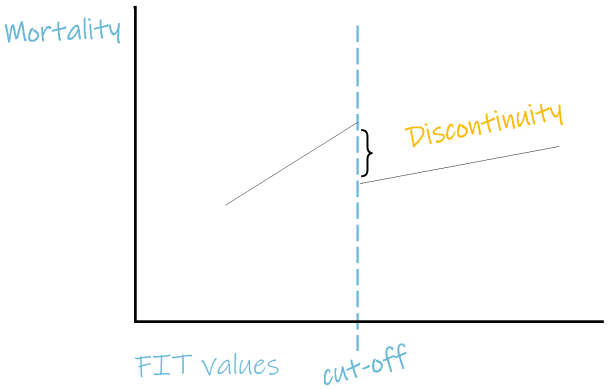

Figure 9 Discontinuity in mortality rate caused by the cut-off of a continuous variable (fHb) prompting intervention

If being referred to colonoscopy has an effect, a *discontinuity* in mortality will appear at the 20 µg fHb/g feces cut-off as those with values immediately above the cut-off have been referred to colonoscopy and those with values immediately below have not. The effect estimated is a *local* effect of the 20 µg fHb/g feces cut-off, i.e. of screening just above vs. below the cut-off.

I calculated survival time for each participant from the date of the FIT result to the date of death, censoring by emigration or end of follow-up (31 December 2019 for CRC mortality and 31 December 2020 for all-cause mortality). Using Cox regression, we estimated a hazard ratio (HR) at the 20 µg fHb/g feces cut-off. We calculated the effect using the following parameterization of the hazard function h:

$$\log\big(h(Y|Z)\big) = \beta_0 + \beta_1(Z-c) + \beta_2 1_{(z\geq c)} + \beta_3(Z-c)1_{(z\geq c)}$$

where Z denotes the fecal hemoglobin values, Y is the outcome, c is the cut-off (20 µg fHb/g feces), β1 is the slope of the line below the cut-off, β1 + β3 is the slope of the line above the threshold and β2 estimates the effect of the screening, as it is the difference of the intercepts of the two slopes and thus the discontinuity (HR) due to being referred to colonoscopy vs. not being referred to colonoscopy.

As an approximation of the absolute effect, I multiplied the observed rates of all-cause and CRC mortality just below the FIT cut-off per 1,000 person years by the estimated HRs. This provided an estimate of the rate of deaths avoided for those screening positive instead of negative. The 1,000 person years were divided by this rate to report the number of persons screening positive per year for each death avoided.

In a sensitivity analysis, I conducted a negative control exposure analysis (105) in which the RDD analysis was repeated for cut-offs expected to produce null results. I applied 15 and 25 µg fHb/g feces cut-offs within a narrow range (12–<18 and 22–<28, respectively). A discontinuity should not be seen at these cut-off values as they do not determine colonoscopy referral.

Results

Results from the four studies are summarized below, and full versions of the results can be found in Paper I-IV.

Study I — Review on mental disorders and participation

We included 17 studies in our scoping review of the literature on mental disorders and participation in CRC screening (Table 2). Of the included studies, 12 were from North America (USA and Canada), four were from Asia (South Korea, China, Japan), one was from Australia and two were from Europe (England and multinational). We found a lack of large studies offering precise estimates. Many of the existing studies were small and/or had a risk of bias from self-reported information on mental disorders and/or screening participation. The majority of the studies were from opportunistic screening programs or programs using the gFOBT stool test. Evidence from organized screening programs and/or universal health systems was scarce (9).

The study is published in Paper I, and we concluded:

> "This scoping review indicates that persons with severe mental illness participate less in CRC screening compared to the background population. For common mental illness, the pattern of participation is less clear and there were few robust studies. The included studies pointed both towards lower and higher participation among persons with depression or depressive symptoms, and the results for persons with anxiety were imprecise."
>
> *MD Jørgensen, EM Mikkelsen, R Erichsen, MK Thomsen: Mental illness and participation in colorectal cancer screening: a scoping review, p. 7. Scandinavian Journal of Gastroenterology 2022* (9).

Table 2 Study characteristics and results of papers in scoping review. Condensed and adapted from Paper I (9).

Author, year, country and study population	Exposure, screening method, adjusted variables	Participation estimates (95% CI if provided)
Descriptive cross-sectional studies		
Friedman et al. 2004, USA **Low-income female psychiatric patients ≥50 years (N=77)**	Self-reported: major depression, bipolar disorder, anxiety, schizophrenic disorder Self-reported: stool test	Participated: 35%
Inagaki et al. 2018, Japan **Patients at a psychiatric medical center (N=97)**	Register diagnosis: schizophrenia diagnosed by current primary psychiatrist	Participated: 13.4% (6.6%;20.2%)

	Register data: FIT	
Xiong et al. 2008, USA **Patients at health clinics (N=68)**	Register diagnosis: unspecified Self-reported: stool test, FS or colonoscopy	Never participated: 56%
Analytic cross-sectional studies		
Browne et al. 2019, USA **Veterans (N=4,461,247)**	Register diagnosis: depression, post-traumatic stress disorder (PTSD), anxiety, serious mental disorders (bipolar or schizophrenia), substance abuse Register data: stool test, flexible sigmoidoscopy (FS) or colonoscopy	No mental disorders: 82.2% Depression: 83.0% PTSD: 85.1% Anxiety: 81.5% Serious mental disorders: 76.0% Substance abuse: 77.0%
Fujiwara et al. 2017, Japan **Patients at a psychiatric medical center (N=224)**	Register diagnosis: schizophrenia Self-reported: FIT Adjusted for age	General population: 40.7% Schizophrenia: 25.1%
Hategekimana et al. 2016, Canada **(N=11,386)**	Self-reported: self-perceived mental health status at baseline Self-reported: gFOBT or FIT depending on state of residency Adjusted for age, gender, race, immigration status and education	Poor mental health: 30.5% ORs for participation, vs. poor mental health: Fair: 1.61 (0.81;3.19) Good: 1.30 (0.72;2.31) Very good: 1.53 (0.86;2.72) Excellent: 1.53 (0.86;2.71)
Kearns et al. 2018, England **Yorkshire Health Study (YHS, N=7,330)** **English Longitudinal Study of Ageing (ELSA, N=6,105)**	Self-reported: depression and anxiety Self-reported: gFOBT Adjusted for age, gender, ethnicity and other chronic conditions	ORs for non-participation vs. persons with no long-term condition (including somatic): Anxiety: YHS 1.13 (0.90;1.42), ELSA 0.95 (0.65;1.40) Depression: YHS 1.40 (1.11;1.76), ELSA 0.98 (0.69;1.41)
Kodl et al. 2010, USA **Veterans 50 to 75 years old (N=855)**	Register diagnosis: unipolar or bipolar depression, PTSD, psychosis, anxiety and substance abuse Self-reported and register data: stool test, FS, double-contrast barium enema or colonoscopy	No mental disorders: 46.9% ORs for participation vs. no mental disorders[1]: Any mental disorders: 0.97 PTSD: 0.69 Depression: 0.43 Psychotic disorder: 0.41

	Adjusted for demographic factors and out-patient visits	Anxiety: 0.54 Substance use: 0.54
Mo et al. 2014, China, Hong Kong **Patients at community mental health services (N=236)**	Self-reported: schizophrenia, depression and bipolar disorder Self-reported: stool test, FS, colonoscopy, double contrast barium enema or CT colonography	Mental disorders: 21.6% ORs for participation vs. persons with schizophrenia: Depression: 3.42 (1.52;7.70) Bipolar disorder: 3.71 (1.01; 17.07)
Peytremann-Bridevaux et al. 2008, Europe (Austria, Denmark, France, Germany, Greece, Italy, the Netherlands, Spain, Sweden, and Switzerland, N=18,560)	Self-reported: depressive symptoms Self-reported: FS or colonoscopy Adjusted for sociodemographics, smoking, alcohol, chronic diseases, function level and country	No depressive symptoms: 16.8% OR for participation for persons with depressive symptoms vs. without: 1.3 (1.1;1.6)
Shin et al. 2020, South Korea **Persons ≥50 years (2006: N=5,170,132, 2015: N=13,202,307)**	Register diagnosis: disability due to mental disorders Register data: stool test	No disabilities: 2006: 16.5%, 2015: 33.8% OR for participation among persons with mental disorders vs. no disabilities: 0.78 (0.77;0,79).
Siantz et al. 2016, USA **Survey respondents ≥50 years with a screening recommendation from their doctor (N=15,355)**	Self-reported: Mental disorders in the past 12 months Self-reported: stool test, FS or colonoscopy [1]Adjusted for sex, age and race/ethnicity [2]The above, health literacy, access, and self-rated health	No mental disorders: 21.7% OR for participation among person with mental disorders vs. without: 0.97 (0.70;1.37)[1] 0.89 (0.63;1.25)[2]
Yee et al. 2011, USA **Female veterans aged 50-65 years (N=606)**	Register diagnosis: anxiety, depression, dissociative symptoms, eating disorders, impulse control disorders, manic symptoms, personality disorders, psychosis, substance abuse, somatoform disorders Register data: stool test, FS or colonoscopy Adjusted for age, service connection, insurance, visits to primary care/health clinics	No mental disorders: 70% OR for participation among persons with mental disorders vs. without: 0.85 (0.56;1.28)

Cohort studies		
Aggarwal et al. 2008, USA (N=79,991)	Self-reported: depressive symptoms Self-reported: stool test, lower endoscopy or double-contrast barium enema Adjusted for sociodemographics, CRC family history, alcohol, comorbidities and having a primary care provider	Overall participation: 53% Risk difference in participation among persons with depressive symptoms vs. without: 0.2%-points (-0.07;1.1)
Bhatia et al. 2021, Canada (N=4,782,718)	Register diagnosis: major mood disorders or psychotic disorders Register data: gFOBT, FS or colonoscopy Adjusted for sex, migration status, income, rural residence, primary care model, care visits, and time of screening launch	HR for participation among persons with mental disorders vs. no chronic conditions: 0.88 (0.87;0.88)
Tuesly et al. 2019, Australia (N=760,058)	Prescription registry: schizophrenia or bipolar disorders Register data: stool test Adjusted for age, gender, state and general practitioner visits	IRR for participation among persons with mental disorders vs. without: 0.90 (0.86;0.94)
Yarborough et al. 2017, USA (N=92,445)	Register diagnosis: psychotic disorder or non-psychotic unipolar depression Register data: gFOBT, FIT, colonoscopy, dual contrast barium enema, or FS	HRs for participation among persons with mental disorders vs. without: Psychotic disorder: 1.05 (0.88;1.25) Depression: 1.18 (1.09;1.28)

Study II

Mental disorders and colorectal cancer screening participation

Overall, 54.6% of men and 63.3% of women participated in CRC screening within 90 days of invitation. We estimated participation differences (PDs) and participation ratios (PRs). Men with mild/moderate mental disorders had 4.4 percentage points (95% CI: 4.1;4.7) and men with severe mental disorders had 13.8 percentage points (95% CI: 13.3;14.3) lower participation than men without mental disorders when adjusted for age, comorbidity, education level and immigration status (PDs, Table 3). On a relative scale, this corresponded to 8% lower and 27% lower participation than men without mental disorders (PRs 0.92 (95% CI: 0.91;0.92) and 0.73 (95% CI: 0.72;0.73), respectively). Even though women had higher overall participation than men, the differences in participation according to mental disorders exposure were similar. Thus, women with mild/moderate mental disorders had 3.8 percentage points (95% CI: 3.6;4.1) and women with severe mental disorders had 15.4 percentage points (95% CI: 14.9;15.8) lower participation than women with no mental disorders.

Table 3 Participation proportions, risk differences and relative risks after 90 days, by mental disorder category

		PD (95% CI)		PR (95% CI)	
Mental disorders	Participation	Crude (95% CI)	Adjusted (95% CI)	Crude (95% CI)	Adjusted (95% CI)
Men					
No (reference)	56.0%	0	0	1	1
Mild /moderate	50.6%	-5.4% (-5.7;-5.0)	-4.4% (-4.7;-4.1)	0.91 (0.90; 0.91)	0.92 (0.91; 0.92)
Severe	37.9%	-18.1% (-18.6;-17.7)	-13.8% (-14.3; -13.3)	0.68 (0.67; 0.69)	0.73 (0.72; 0.73)
Women					
No (reference)	65.1%	0	0	1	1
Mild /moderate	60.5%	-4.7% (-4.9;-4.4)	-3.8% (-4.1;-3.6)	0.93 (0.93; 0.93)	0.94 (0.94; 0.94)
Severe	46.9%	-18.2% (-18.6;-17.8)	-15.4% (-15.8; -14.9)	0.72 (0.72; 0.73)	0.75 (0.75; 0.76)

Adjusted for: Age group, comorbidity, education, being an immigrant

When looking at subtypes of severe mental illness, the lowest participation proportions were observed among persons with schizophrenic and psychotic disorders among both men and women. For men, participation was 17.6 to 19.5 percentage points lower among persons with schizophrenic disorders or depression and 10.1 to 12.3 percentage points lower among persons with other severe mental disorders than among men with no mental disorders. For women, the estimates ranged from PD=-19.2% (95% CI: -19.9;-18.6) for persons with schizophrenic disorders to PD=-5.8% (95% CI: -11.9;0.3) for anxiety and 1≤ other severe mental disorder (see further details in Paper II).

Trajectories through the screening program were also affected by mental disorders (Figure 10). Among persons who participated within 90 days, the proportion of positive FITs was 6.6% among persons with no mental disorders, 7.7% among persons with mild/moderate mental disorders and 8.9% among persons with severe mental disorders. Likewise, more tests were non-analyzable among persons with mental disorders (0.14% and for mild/moderate and 0.24% for severe) than among persons with no mental disorders (0.08%).

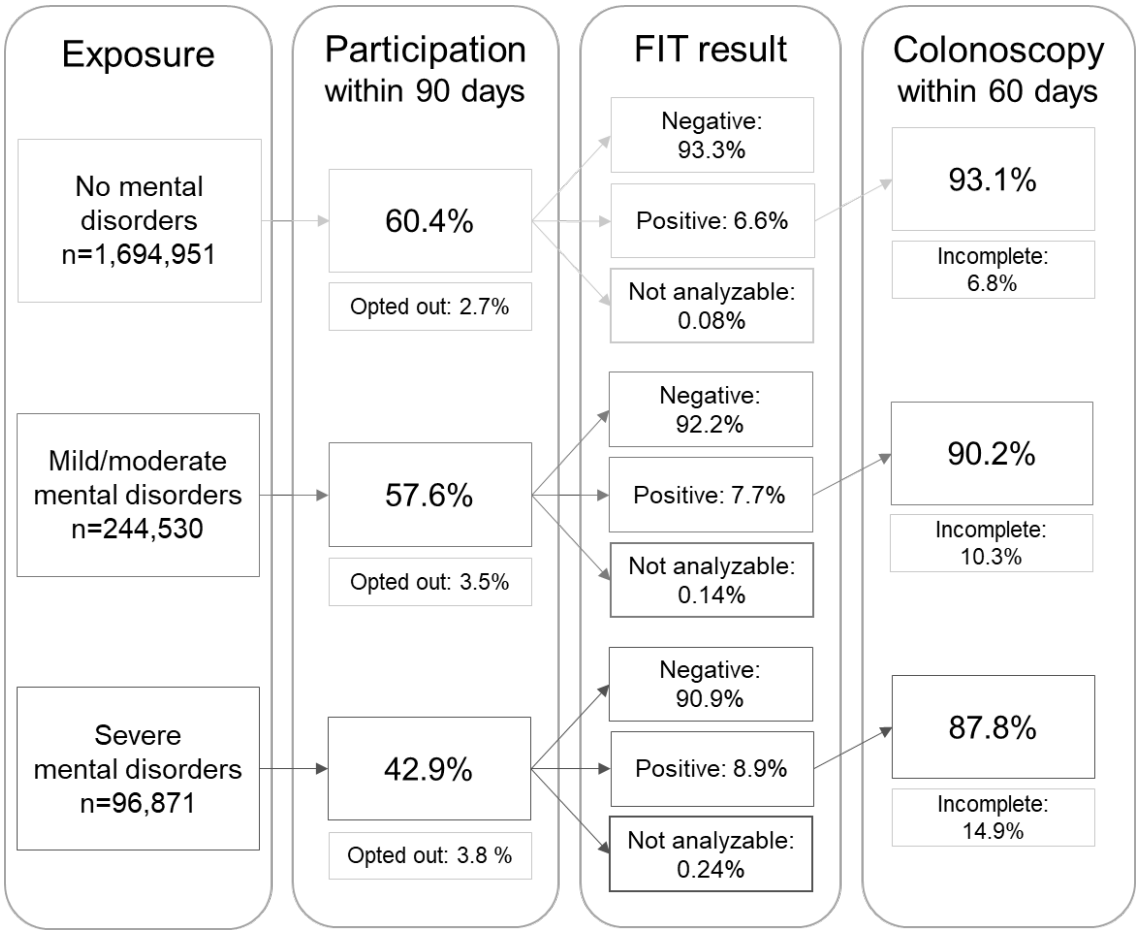

Figure 10 Trajectories throughout the screening process for persons with versus without mental disorders

Persons with mild/moderate and severe mental disorders had lower adherence to colonoscopy (90.2% and 87.8%, respectively, vs. 93.1%), and their colonoscopies were more often incomplete (10.3% and 14.9%, respectively, vs. 6.8%) due to either poor bowel emptying before the procedure or pain, unpassable turns, etc., than among with persons with no mental disorders. The degree of colonoscopy completeness was also less often registered for persons with mental disorders (7.7% and 8.3%, respectively) than for persons with no mental disorders (6.6%).

Study III

Predicting colorectal cancer for risk-stratified screening

In Study III, I developed prediction models for CRC and advanced neoplasia including age, gender and fecal hemoglobin (106). Women had lower odds of colorectal cancer than men in the crude analysis, but not in the full model which included all three variables (Table 4). When combining medium-risk and high-risk adenomas with CRC in the outcome *advanced neoplasia*, women had considerably lower odds than men (full model OR=0.64 (0.61;0.67)). Age had a linear relationship with CRC and advanced neoplasia. Fecal hemoglobin strongly predicted both CRC and advanced neoplasia (106).

Table 4 Logistic regression models predicting colorectal cancer or advanced neoplasia (106).

	OR for colorectal cancer (95% CI)		OR for advanced neoplasia (95% CI)	
	Crude	**Full model**	**Crude**	**Full model**
Gender				
Men (ref.)	1	1	1	1
Women	0.83 (0.76;0.91)	0.96 (0.87;1.06)	0.60 (0.58;0.63)	0.64 (0.61;0.67)
Age (continuous)	1.05 (1.05;1.06)	1.05 (1.05;1.06)	1.03 (1.03;1.04)	1.03 (1.03;1.03)
Fecal hemoglobin, µg fHb/g feces				
20-29 (ref.)	1	1	1	1
30-44	1.37 (1.09;1.73)	1.36 (1.08;1.72)	1.35 (1.25;1.46)	1.34 (1.24;1.45)
45-79	1.86 (1.49;2.31)	1.83 (1.43;2.28)	1.65 (1.53;1.78)	1.64 (1.52;1.77)
80-199	3.26 (2.66;4.00)	3.23 (2.62;3.98)	2.33 (2.16;2.51)	2.32 (2.15;2.51)
200-	9.17 (7.61;11.06)	9.08 (7.51;11.00)	4.32 (4.01;4.64)	4.25 (3.95;4.58)
Constant[a]	-	0.01 (0.01;0.01)	-	0.22 (0.20;0.24)

[a]Age centered at 50 years. Abbreviations: fHb= fecal hemoglobin, OR=odds ratio, ref=reference.

The model predicting CRC had discriminatory performance with an AUC=75% (95% CI:74;76), and the model predicting advanced neoplasia had an AUC=67% (95% CI:66;68), Figure 11. The models had similar AUCs in temporal validation and were well calibrated (106).

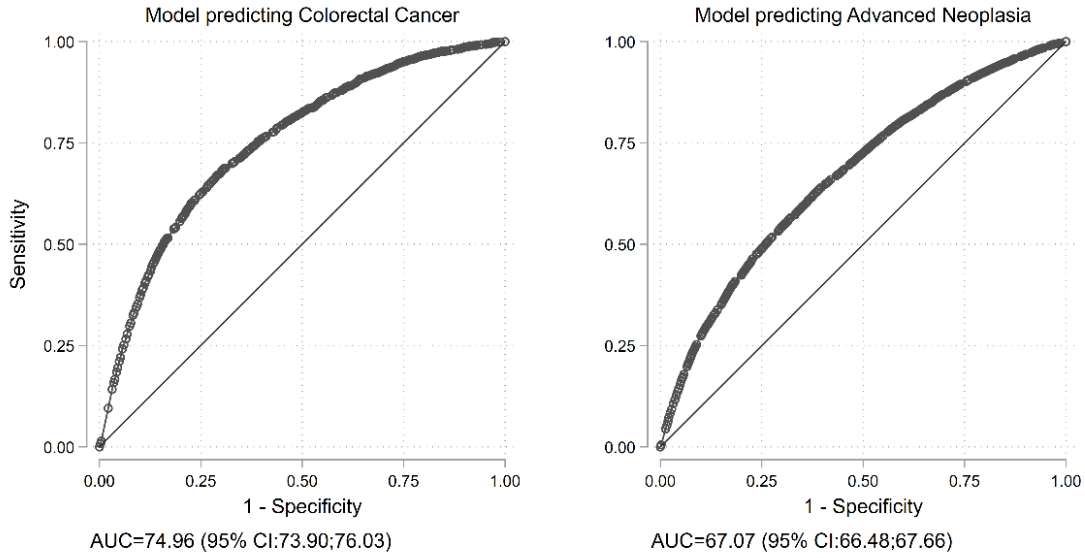

Figure 11 Receiver-operator-characteristic (ROC) curves for models predicting colorectal cancer and advanced neoplasia (development dataset) (106).

I conducted the comparison using a risk cut-off to select participants to undergo colonoscopy using FIT alone. The proportion of positive FITs expected by health authorities at initiation of the screening program (5.2% (103)) corresponded to an approximate 24% reduction in the number of colonoscopies, equal to a FIT cut-off at 33.4 µg fHb/g feces. This was used as the reference.

Table 5 Comparison of colonoscopy findings using FIT cut-offs vs risk cut-offs per 10,000 individuals with a FIT ≥20 µg Hb/g feces and colonoscopy (106).

Colonoscopies		**CRC**		**Medium/high-risk adenomas**		**Low-risk adenomas**	
	n	n	Increase, n Ratio (95% CI)	n	Increase, n Ratio (95% CI)	n	Increase, n Ratio (95% CI)
FIT cut-off, 33.4 µg Hb/g	7,555	548	ref.	2,429	ref.	908	ref. 1
Risk cut-off, CRC[a]	7,554	566	↑18 1.03 (1.02;1.05)	2,453	↑24 1.01 (1.01;1.01)	929	↑21 1.02 (1.01;1.03)
Risk-cut-off, AN[a]	7,577	558	↑10 1.02 (1.01;1.03)	2,508	↑79 1.03 (1.03;1.04)	955	↑47 1.05 (1.04;1.07)

[a]Medium-risk and high-risk adenomas. [b]Logistic regression models predicting risk of CRC or AN from age, gender and fecal hemoglobin. Cut-off CRC risk 2.13475%, and AN risk 24.12302%.
Abbreviations: AN=advanced neoplasia (medium/high-risk adenomas combined with CRC), CRC= colorectal cancer.

Compared with the 33.4 µg fHb/g feces FIT cut-off, the CRC risk cut-off identified 18 more CRCs and 24 more medium/high-risk adenomas (Table 5). This is a relative increase of 1.03 (95% CI:

1.02;1.05) more CRCs and 1.01 (95% CI: 1.01;1.01) more advanced neoplasia for a similar number of colonoscopies. An additional 21 low-risk adenomas would be identified; a relative increase of 1.02 (1.01;1.03). The model predicting advanced neoplasia identified 10 more CRCs, 79 more medium/high-risk adenomas and 47 more low-risk adenomas than the FIT cut-off (106).

Study IV

Effect of FIT screening on mortality

In Study IV, we used the RDD to estimate the effect of FIT screening on mortality. The study included 16,428 persons to estimate all-cause mortality in the narrow fecal hemoglobin range. Median observed person time was 4.4 years, with maximum 6.8 years. During follow-up, 1,140 deaths occurred. The distribution of study population characteristics was similar across the 20 µg fHb/g feces cut-off in both ranges, as is depicted in Figure 12.

The HR comparing all-cause mortality immediately above with all-cause mortality immediately below the cut-off was 0.87 (95% CI: 0.69;1.10); estimated from the narrow fecal hemoglobin range. The effect on all-cause mortality was similar in the wide range, but with a slightly more precise CI (HR=0.88 (95% CI: 0.75;1.07)) due to more than double the number of deaths (n=2,363).

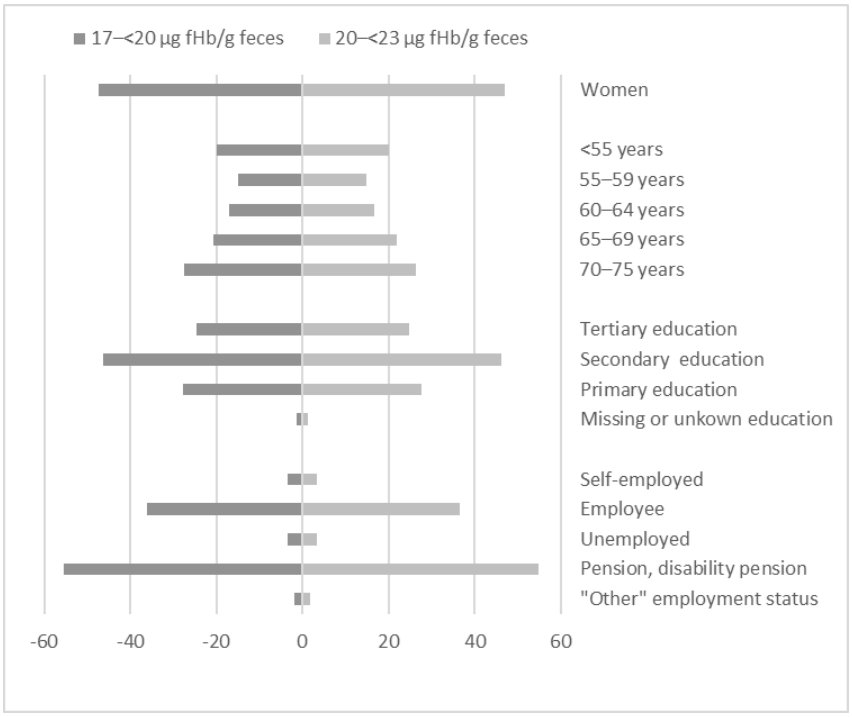

Figure 12 Proportions (%) of study population characteristics below versus above the 20 µg fHb/g feces cut-off, narrow range all-cause mortality study population.

For CRC mortality, too few deaths occurred in the narrow range, but the HR was 0.49 (95% CI: 0.17;1.41) in the wide range. Figure 13 and Figure 14 show these regression functions graphically

with 95% CIs (dotted lines). The bold dashed line shows the ≥20 µg/g feces cut-off, and the light dashed lines show data ranges to fit the model.

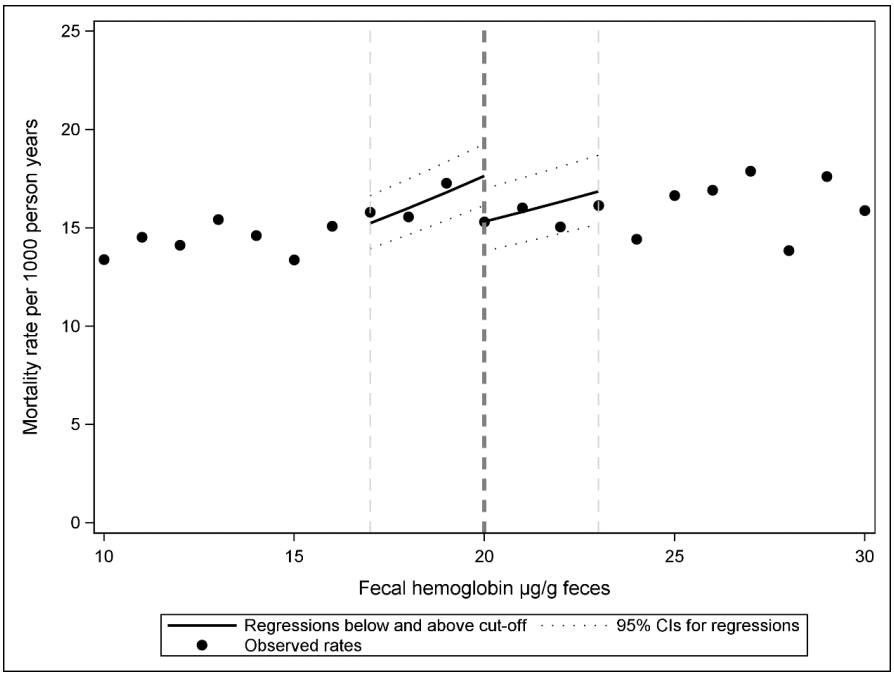

Figure 13 Regression discontinuity plot of all-cause mortality by fecal hemoglobin value, narrow range.

The discontinuity in the hazard functions appears at the 20 µg fHb/g feces cut-off. The functions were fitted to all fecal hemoglobin data points within each range; however, the filled circles depict mortality rates per 1,000 person years calculated for each fecal hemoglobin value rounded down to nearest integer and are included for comparison.

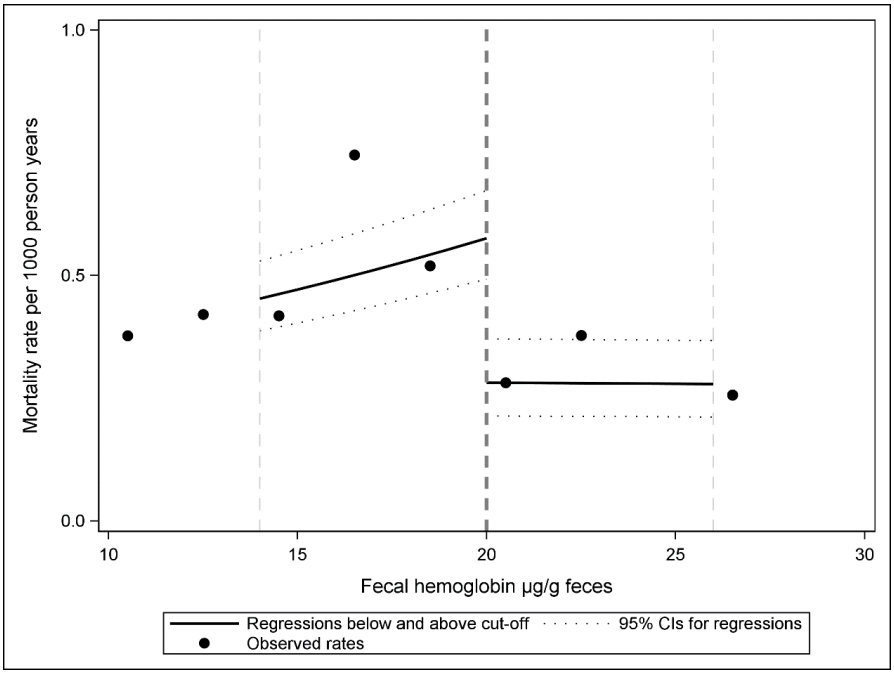

Figure 14 Regression discontinuity plot of CRC mortality by fecal hemoglobin value.

At 19 µg fHb/g feces (rounded), just below the cut-off, observed rates of death were 17.3 per 1,000 person years for all-cause mortality and 0.52 per 1,000 person years for CRC mortality. As an approximation of the absolute effect, I estimated that one death of all causes per 444 persons per year and one CRC death per 4,000 persons per year were avoided for those referred to colonoscopy as they screened positive instead of negative.

In sensitivity analysis, analytically changing the cut-off to 15 and 25 µg fHb/g feces led to less well balanced study population characteristics across the cut-off. Persons above the cut-offs were slightly more often men, in the oldest age groups, had a tertiary education and lower health service utilization than those below the cut-off. The HRs for all-cause mortality were 0.96 (95% CI: 0.79;1.17) for the 15 cut-off and 1.13 (95% CI: 0.86;1.50) for the 25 cut-off . The analysis for CRC mortality included 60 and 15 deaths for the 15 and 25 cut-offs, respectively, with HRs of 1.51 (95% CI: 0.52;4.38) and 0.56 (95% CI: 0.06;4.84), respectively.

Discussion

Main findings

In Study I, we reviewed the existing literature on the participation of persons with mental disorders in CRC screening. The identified studies were predominantly from opportunistic screening programs in the USA and Asian countries. All included studies found lower participation among persons with severe mental disorders; however, the pattern was less clear for persons with mild/moderate mental disorders. Gaps in current evidence include lack of large e.g. register-based and/or population-based studies, not sufficiently precise results, and few studies have been conducted on FIT-based screening programs and organized as opposed to opportunistic screening programs (9).

In Study II, we compared the participation of persons with and without mental disorders in the Danish CRC screening program. Overall, participation was 54.6% among men and 63.3% among women. Men and women with mild/moderate mental disorders and severe mental disorders had lower participation than men and women without mental disorders. Persons with schizophrenic disorders had the lowest participation in CRC screening, but participation was lower for all categories of mental disorders than for persons with no mental disorders. Trajectories throughout the screening program were also affected; compared with persons with no mental disorders, persons with mental disorders had a higher proportion of both positive and non-analyzable FITs as well as a lower adherence to colonoscopy, and their colonoscopies were more often incomplete.

In Study III, we developed models predicting the risk of CRC or advanced neoplasia from data on age, gender and fecal hemoglobin values. The CRC model identified more CRCs than the advanced neoplasia model, and the advanced neoplasia model identified more adenomas than the CRC model. With existing data, risk-stratified FIT screening using a risk cut-off instead of a FIT cut-off may slightly improve selection to colonoscopy of those at highest risk of cancer and adenoma (106).

In Study IV, we estimated the local effect of being referred to colonoscopy for persons with fecal hemoglobin values around the cut-off for colonoscopy. The estimated effect was a HR=0.87 (95% CI: 0.69;1.10) for all-cause mortality and a HR=0.49 (95% CI: 0.17;1.41) for CRC mortality. Both estimates have broad CIs; and especially for CRC mortality, the number of outcomes for persons in the relevant fecal hemoglobin ranges was limited. Precision may increase with longer follow-up time and more screening rounds.

Results in the context of other studies

Study I+II

Mental disorders and participation in colorectal cancer screening

Compared with the results summarized in Study I and the results in the review by Solmi et al. (60), we were able to provide more precise estimates in Study II due to a large sample size. Like the studies included in Study I, we also found lower participation in CRC screening among persons with severe mental disorders; however, various effect measures were used, and we therefore cannot perform a direct comparison of estimates. We found lower participation for all subtypes of hospital-diagnosed mental disorders and for disorders such as depression and anxiety which had ambiguous results in some other studies (9). For persons with common mental disorders, our scoping review reported ambiguous results; the review by Solmi et al. reported no association (9, 60). In Study II, however, we found lower participation among persons with mild/moderate mental disorders than among persons with no mental disorders. Our definition of mild/moderate mental disorders relied on treatment, whereas other studies used register-based diagnoses or self-reported information (9).

Study III

Risk-stratification in fecal immunochemical test screening

Shortly after the publication of Study III, a scoping review of risk-stratified CRC screening was published by Cairns et al. (107). This review found 13 studies examining the diagnostic findings using a risk-stratified method compared with FIT alone. They found overall that "Risk models do appear to show promise in refining existing risk stratification guidelines but most were not externally validated and less than half achieved good discriminatory power" (107, p. 1). Less than 40% of the included studies reported discrimination for their prediction model. Measures of discrimination and calibration are needed to assess the validity of a prediction model and to enable comparisons between models (69).

I found only one study, by Cooper et al., directly comparable to Study III. Like our study, they derived a prediction model from information available in the screening program using data from persons with a positive FIT, and they evaluated performance according to both discrimination and calibration (71). Like our study, they concluded that risk-stratified screening was superior to the FIT cut-off in selecting persons for colonoscopy. However, they predicted the risk of advanced neoplasia and included former screening participation as a fourth variable in their prediction model. Their model had an AUC of 66%; similar to our model in predicting advanced adenomas (67%).

A study from the Netherlands was not directly comparable to ours as all persons underwent colonoscopy irrespective of their FIT result. The study included 1,011 persons, and the derived model

predicting advanced neoplasia was not superior to FIT alone as AUCs were 0.71 (95% CI: 0.65;0.78) for the risk model vs 0.69 (95% CI: 0.63;0.75) for FIT alone (108).

The predictive performance of the models may be improved through inclusion of additional predictors. However, others found that neither self-reported data on family history of CRC (109) nor lifestyle factors (73) improved the AUCs of their prediction models. On the other hand, models including clinical data from health care records reported a high AUC of 86% (110), and a model including fecal microRNA had an AUC of 90% (111).

Study IV

The effect of fecal immunochemical test screening on mortality

The RDD was applied by Kadiyala & Strumpf to estimate screening coverage and CRC detection at the threshold starting age for screening programs in the USA and Canada (112, 113), but no other study has estimated the effect of FIT screening on mortality using the RDD. As the effect estimated using this design is local, i.e. it falls around the cut-off for colonoscopy referral, only results from studies using the same design could be interpreted in a similar way. In addition, previous observational studies using other designs reported RRs, whereas we estimated a HR. The two effect measures do not have the same interpretation, and they are not comparable in magnitude. If the RR is the reduction in the cumulated risk of death after follow-up until time t, the HR is the reduction in the instantaneous risk of death at time *t*, given that the person survived until time *t* (114).

Thus, the existing literature offers no estimates directly comparable to ours for all-cause mortality (HR=0.87 (95% CI: 0.69;1.10)). Most closely related to my study is Njor et al.'s comparison of persons invited vs. not-yet-invited in the beginning of the Danish screening program. This study reported a protective effect of screening on CRC mortality (RR=0.83 (95% CI: 0.66;1.03)). However, this study design has a major limitation. Unlike the RDD, it cannot be reapplied later in the screening program as the random sequence of invitation according to birth months occurred only in the first screening round in 2014-2017. The design is therefore specific to the implementation screening round, and follow-up was stopped at the time of invitation for the control group. This limited follow-up to less than four years (median 3.3 years) limits the design to evaluation of the prevalence round (i.e. a population with predominantly prevalent cancers).

The four RCTs on gFOBT had an effect (RR=0.87 (95% CI: 0.82;0.92)) on CRC mortality similar to that reported in Njor et al.; however, in contrast to Study IV, they found no effect on all-cause mortality.

In general, it is a challenge for observational studies on screening to identify suitable comparison groups, except in the beginning of a screening program when random invitation sequences or staggered geographical implementation can be studied. Therefore, the application of the RDD to evaluate the effect of FIT-based screening on mortality is valuable and novel. It allows for causal interpretation due to its quasi-experimental design, and it can be repeated at various time points after program implementation.

The RDD makes the contrast between the 'intervention group' and the 'control group' clearer than would be possible in RCTs or comparisons of invited vs. not-yet-invited participants as our study included only participants. When comparing invited with not-invited (or not-yet-invited) participants, a rather small proportion of the study population has the possibility of experiencing an effect of screening. Thus, the intervention group is diluted with non-participants and persons with negative FITs, and neither are referred to colonoscopy. In our study, 90% of those with a positive FIT underwent colonoscopy within a few months (81). This may partly explain why we were able to estimate a local effect of FIT screening on not only CRC mortality but also all-cause mortality.

Methodological considerations

While the goal of epidemiologic studies is a precise and valid estimation of effects, no study can be perfect, and estimates come with some degree of error. This error can be divided into random error (precision) and systematic error (validity). Systematic error, or bias, can arise from selection, information or confounding issues (115), in addition to screening-specific biases (22). Study I was a scoping review, and Study III was a prediction study, which did not seek to estimate causal effects. The quality of the scoping review and the validity of the prediction study are therefore discussed separately.

Random error

To some degree, studying nationwide screening programs comes with the gift of precision as random error generally decreases with increasing sample size. This is especially evident in Study II on participation, where the study population was more than two million persons invited to screening in 2014-2018, and the CIs were narrow. Study III also has a reasonable sample size as far as the number of variables included in the prediction model is concerned, whereas interpretation of the results from Study IV is challenged by moderate to low precision. Despite all the virtues of the RDD applied in Study IV, the sample size is restricted because we analyzed hemoglobin values close to the cut-off. Combined with the marginally sufficient follow-up time available to us (median 4.4. years), estimates

of especially CRC mortality were imprecise with CIs ranging from a highly protective effect (95% CI lower bound=0.17) to a harmful effect (95% CI upper bound=1.41). Our estimate of HR=0.49 is, however, our best suggestion, but this needs to be re-estimated in the Danish program using longer follow-up time and to be investigated with the same design in other FIT programs.

Study I

Quality of the scoping review

Despite broad literature searches in several databases, we can have missed papers relevant to the review question in Study I. We identified more papers than the review by Solmi et al.; however, they included six studies that we did not (60). The six papers were all from the USA and concerned the quality of primary care provided to persons with mental disorders (116–121). Thus, none of the papers had the word "screening" in their titles. However, these studies provided little new information on participation in CRC screening for persons with mental disorders. In the USA, regular health checks and screening for several cancers and other conditions are handled by primary care practices and not national screening programs (117). This system may entail different mechanisms of screening participation than those seen in the Danish screening program in Study II, and studies may not be comparable across systems. The various designs of CRC screening programs across countries makes interpretation of the meta-analysis by Solmi et al. difficult to interpret, which underlines the usefulness of the scoping review design to map existing evidence.

Selection

The selection of persons into a study and factors affecting participation (including continued participation) in a study can cause bias. Bias may arise in situations where selection leads to different associations between exposure and disease among persons in the study than among persons eligible for the study (122), i.e. if both the studied exposure and outcome affect inclusion or participation (123). Because we used high quality national registry data with linkage between data sources in Study II-IV, all eligible persons were included and accounted for. Thus, we consider the overall the risk of selection bias to be small (124).

Study II

Informative censoring

In Study II on mental disorders and screening participation, we designed the study to avoid selection bias due to informative censoring (125). Some studies on participation in screening excluded persons who died in the follow-up period (45, 90, 91). However, as it is well documented that persons with mental disorders have high excess mortality (92, 93), we designed the study as a cohort study and used a regression model while taking censoring from emigration and competing risk

from death into account. If (lack of) participation up until a person's death (or emigration) is ignored, this could distort the results, either towards or away from the null, depending on their likelihood of participation relative to persons included in the study.

Information

Problems with the data collected in a study can lead to information bias, often called misclassification of categorical variables (122). Information bias can arise in two distinct ways (126). The first is through information not correctly recorded; errors could, e.g., be measurement error or they could be due to respondent recall error, misreporting and wrongful coding in databases. The second way misclassification can arise is by defining a study variable in a way that is different from the "true" definition (126).

Study II-IV were conducted using data from existing databases and registers. Thus, there is no risk of error due to respondent recall error. However, the data used were not recorded primarily for research purposes. The screening database and the colorectal cancer database are maintained for quality monitoring, whereas all the other data sources are administrative registries used for reimbursement, public planning, etc. This may affect how some diagnosis and procedure codes are reported, e.g. in situations where several different codes could be relevant, a reimbursement incentive may favor one over the other (39). However, the Danish national registers are generally of high validity (39). Residents are prospectively registered at almost all points of contact with the healthcare and public systems with a high degree of completeness (39). Yet, some information problems may exist with the variables used in Study II-IV.

Study II

Definition of mental disorders

In Study II, our definition of mental disorders distinguished between severe mental disorders and mild/moderate mental disorders according to our findings in Study I. In Denmark, severe mental disorders are treated at hospitals, while mild and moderate mental disorders are generally treated by general practitioners, practicing psychiatrists and psychologists (127). First, we hypothesized that mild/moderate mental disorders would also be treated at out-patient hospital visits. Therefore, we expected that a group of patients would have a diagnosis of mild depression, anxiety or similar disorders recorded at out-patient visits with no in-hospital stays. However, data revealed that this was not the case as only a few hundred persons were in this category. It appeared that out-patient visits for the included diagnoses most often took place after or in connection with hospitalization, therefore constituting severe mental disorders in our definition.

Some misclassification between mild/moderate and severe mental disorders may remain, but its degree and direction depend on what can be considered the true definition of severity. Any misclassification of the mental disorder variable could not have been affected by the outcome, screening participation. Therefore any misclassification would not be differential, i.e. it would be independent of the value of the outcome. For a dichotomous variable, non-differential misclassification of the exposure on average causes bias toward the null (126). Our mental disorders variable was not dichotomous, however; and therefore the middle category of mild/moderate mental disorders could be biased away from the null, if it had been mixed with persons who should be in the severe category (126).

We used a look-back period of five years, which is to some degree an arbitrary duration. This time period was chosen to identify pertinent mental disorders able to affect screening participation at the time of invitation. However, treatment of mental disorders takes time, and the influence of the disorders on daily life is likely to carry on for some time after the last health care record; we therefore did not use a shorter look-back period. We used register data and thus avoided bias due to self-report and errors of information recall, but some other studies measured self-reported mental health or used clinical interview tools (9).

Study III

Fecal hemoglobin

The fecal hemoglobin value was used as a predictor in Study III and to model the effect of colonoscopy referral on mortality in Study IV. The quantitative measurement of fecal hemoglobin may be encumbered with some measurement error; however, this is most likely a source of "noise" or a random error and not an instance of misclassification. A specific apparatus could hypothetically be miss-calibrated, leading to a systematic measurement error; however, this would be non-differential, i.e. the same for all analysis results from that apparatus and independent of the outcome. Another issue related to FIT measurements of fecal hemoglobin, is that values above 200 µg/g feces are truncated in the Danish CRC screening database and recorded as the value 200 (106). Therefore, I had to categorize the fecal hemoglobin value in Study III instead of using it continuously as is otherwise advised to optimize prediction (69).

Study IV

Colorectal cancer mortality

Cause-of-death data are considered of suboptimal quality due to uncertainties in determining the underlying cause of death (86). We sought to address this in Study IV, by including all levels of codes in the cause of death registry for the definition of CRC death. Including all levels

of codes will increase the ability of the cause of death registry to identify death from CRC. However, it may also increase the number of false positives, i.e. cases where CRC, although present, in fact did not contribute to death. This would, in turn, decrease specificity as specificity is the number of true negatives divided by the sum of true negatives and false positives. In addition, if there is misclassification of CRC death, it is most likely differential as it may to some degree depend on the exposure (fecal hemoglobin). Persons with a positive FIT are more likely to have a prevalent CRC detected due to colonoscopy referral. A person with a negative FIT could have an (undiagnosed) CRC that contributed to death, and this persons would less likely be registered as dying from CRC than if they had received colonoscopy. This misclassification would underestimate CRC mortality below the cut-off and therefore bias the estimates towards the null.

Confounding

Confounding is the confusion of effects. In observational studies (and experimental studies where randomization did not succeed), confounding can occur because the exposed group differs from the unexposed group on other conditions than the exposure itself affecting the risk of the outcome (122). A confounder is thus a condition, which co-occurs with or causes the exposure studied, and also affects the risk of the outcome. Confounders can be adjusted for either in the analysis or the study design, to prevent them from confounding the estimates (128). I adjusted for confounding in Study II, and avoided confounding by design in Study IV.

Study II

Mental disorders and socioeconomic position

When estimating the association between mental disorders and participation in CRC screening, I based the variable set used for confounder adjustment on a DAG of assumptions. This could leave residual confounding of the estimates from unknown and unmeasured confounders. Results attenuated slightly after adjustment for age, education level and comorbidity score in the main analysis. However, it is also important not to overadjust estimates, e.g. by adjusting for mediators and thus close paths through which mental disorders affect participation. Results attenuated further after adjustment for living alone instead of comorbidity score; however, it may be more likely that mental disorders lead to living alone (especially for severe mental disorders) than the other way around, and thus adjusting for living alone removes the effect of mental disorders that goes through living alone.

Study IV

Randomization and healthy screenee bias

The quasi-experimental design of Study IV should ensure absence of confounding. The balanced distribution of covariates observed strengthens the assumption that measurement

variation acted as randomization in this case. However, this also entails assuming a balanced distribution of any unobserved confounders, something that naturally cannot be confirmed.

Healthy screenee bias arises because persons with generally healthy behavior are more likely to take up preventive services such as screening. This was demonstrated in Study II where both persons with mental and somatic morbidity participated less in screening than did persons without such morbidity. Self-selection into a study has traditionally been considered a selection bias; however, it is more accurately a type of confounding (122). If the degree of healthiness or preferences for healthy behavior could be compared between participants and non-participants, this effect could be adjusted for in the analysis. Several correction factors for healthy screenee bias have been proposed (129) and, e.g., Njor et al. applied a correction (the ratio between non-participants and the not-yet invited persons) when comparing participants and non-participants among the invited in a secondary analysis (55).

Healthy screenee bias is avoided in intention-to-treat comparisons of invited vs. not-invited to FIT, e.g., in a randomized trial. The RDD in Study IV also prevented this bias as all persons in the analysis were participants who all self-selected into the screening program.

Screening specific biases

As described in the background section, lead-time and length-time bias can occur in studies involving screening when persons with screen-detected CRC are compared with persons with CRC detected due to symptoms. No such comparisons were made in the studies of this thesis, and lead-time and length-time bias were therefore avoided.

Study III

Validity in prediction studies

The goal of a prediction study is to provide a valid *model* for prediction. Therefore, it is important to have a well-specified model to be able to evaluate its performance and to validate the model in an external population if it is intended for general use (69, 98). Thus, I described the distribution of age, gender and fecal hemoglobin values, addressed their linear fit, evaluated the performance of the models in terms of discrimination, calibration and findings compared with a FIT cut-off, and validated the model temporally(106). Calibration showed that the model did not systematically overestimate or underestimate risk, and discrimination was moderate.

Applying another statistical model or including more or potentially stronger predictors may increase the ability of the model to discriminate between those with and without CRC or advanced adenomas. External validation could also be advisable to increase the quality of the prediction study. However,

due to the population, design and prioritization differences between screening programs, it may not be beneficial to seek to establish one universal model of risk stratification for FIT screening programs (106).

Main conclusions

This thesis contributed to the international body of evidence relevant for the decision to screen for colorectal cancer. We found no evidence to weaken the resolve to screen for colorectal cancer using FIT. While Study I, II and III pointed to potentials for optimization, Study IV found a beneficial effect of the screening program.

Study I

The existing literature on CRC screening among persons with mental disorders uniformly showed that persons with severe mental disorders such as schizophrenia or major depression participated less in CRC screening. However, the results from studies of more common mental disorders were ambiguous, and few studies addressed this issue in population-based or FIT-based screening programs (9).

Study II

Persons with severe but also persons with mild/moderate mental disorders participated less in the Danish CRC screening than did persons without mental disorders. This difference persisted after adjustment for socioeconomic conditions; thus, mental disorders constitute a separate barrier towards CRC screening participation.

Study III

The models predicting CRC or advanced adenomas from age, gender and fecal hemoglobin values were superior to FIT alone in identifying cancers and adenomas, and their use reduced the number of unnecessary colonoscopies (106).

Study IV

Persons who had a fecal hemoglobin value just above the FIT cut-off, and who were therefore referred to colonoscopy, had lower all-cause mortality than those whose values were just below the cut-off. The analysis of CRC mortality also showed a reduced mortality due to colonoscopy referral; however, the results were encumbered with considerable imprecision.

Perspectives and implications

The Danish CRC screening program is now fully implemented; yet, only 721 out of 2,655 incident cases of CRC in 2020 were detected through screening (37, 130). More than 800,000 persons were invited that year, and 21,500 colonoscopies were conducted within the screening program (37). Screening of an average risk population involves a considerable number of persons and resources, both in the target population and the healthcare system. This underlines the need for continuous evaluation of the implemented screening program to monitor the appropriateness of continuing screening. Below, I provide some perspectives and possible implications regarding each of the aspects studied.

Equal access to screening participation

Study I+II

Denmark has a social-democratic welfare state characterized by universal and relatively generous state provision of health care (94). It has been a puzzle within social epidemiology why the Scandinavian social democratic countries do not have the lowest levels of health inequality. However, Bambra points to several possible explanations for this, including how primarily the middle classes take up universal health messages and health-promoting interventions. Thus, while universal health interventions improve the health of the population on average, it may improve health status more for those already healthiest, creating *intervention-generated health inequalities* (94). Population cancer screening programs could be examples of such inequality-generating interventions. The entire population within the age range is invited to screening; however, as shown in Study II, participation is higher among women, persons who are educated, have sufficient economic resources, live with others, and do not suffer from somatic diseases or mental disorders.

The principles of screening include that the screening test itself should be acceptable to the target population, that all screening program components should be socially acceptable to screening participants and that the screening program should be integrated within the broader healthcare system (3). Screening programs can be adjusted to fit not only the well-off but also groups with low participation today. The literature review in Study I found studies reporting a higher increase in participation among persons with mental disorders than among persons without such disorders after the transition from gFOBT to FIT (a more acceptable test), and several studies reported higher participation among persons with mental disorders if they were registered with a GP or if their GP recommended screening (related to integration with the broader healthcare system) (9).

We demonstrated the inequality in participation between persons with and without mental disorders in Study II. A recent study from Japan, evaluating the feasibility of a case management approach to support participation in CRC screening among persons with schizophrenia, reported that the majority of participants found such an approach acceptable. The intervention component most often reported as helpful was in-person counselling on the screening process conducted by a psychiatric care provider (131). Future research could explore offering increased support for screening decisions targeting persons with mental disorders, e.g. within psychiatric health care and by general practitioners.

Study III

Risk-stratified screening to prioritize colonoscopy

Risk stratification can improve the performance of a screening test and the overall benefit of CRC screening programs by identifying more persons with relevant findings (CRC and adenomas) and testing fewer false positive. Several studies have sought to develop risk stratification models, including Study III of this thesis. However, Cairns et al. found a huge gap between theoretical models and the practical implementation of risk stratification in CRC screening programs. Few pilot studies of implementation of risk stratification in screening programs exist (107). Due to possible implementation challenges, it may be advisable to only include few and readily available types of information in the stratification method, like age, gender and fecal hemoglobin values included in our prediction model.

Currently, most CRC screening programs only stratify by age by offering screening to a specific age group. Men have a higher background risk of CRC but also a higher tendency for blood in the stool than do women. Clark et al. pointed out that the lower proportion of positive tests for women in programs with a uniform FIT cut-off leads to poorer outcomes for women, including a higher rate of interval cancers, i.e. cancers missed by FIT and detected due to symptoms arising before the next screening, and a lower reduction in CRC mortality due to screening. The FIT cut-offs were stratified by gender in the Swedish and Finnish screening programs with higher cut-offs for men and lower ones for women (132). However, when stratifying FIT cut-offs for men and women, a decision has to be made as to which kind of equality is sought. Should, e.g., the proportion of positives, rates of interval cancers, sensitivities or specificities be equal for men and women? Further development of risk prediction models may prove useful as they allow for more detailed modelling of risk from several parameters (133).

Study IV

Continued evaluation and monitoring of the effect of screening

The ultimate goal of screening is a lower overall mortality, i.e. prolonged duration of life. Related to the principles of screening, the benefit of screening should be documented before program implementation. However, evidence for the benefit of FIT screening is limited and to some degree circumstantial as it rests on gFOBT RCTs from the eighties and nineties. Novel observational study designs are needed to fill this gap and provide direct evidence for the effect of FIT screening.

A recent study proposed that the benefit of *breast cancer* screening is diminishing as developments in treatment have improved survival in all cancer stages, making early detection by screening less crucial (134). Thus, continued evaluation and monitoring of screening benefits is needed to ascertain the remained relevance of CRC screening in the future.

The RDD method enabled estimation of a causal effect of screening, something that is seldom possible in observational data. We encourage repeating the study later in the Danish CRC screening program, and using the RDD to evaluate other CRC screening programs, especially those with longer follow-up and/or even larger target populations than the ones available to us.

References

1. Croswell JM, Ransohoff DF, Kramer BS. Principles of Cancer Screening: Lessons From History and Study Design Issues. Semin Oncol. 2010; 37(3):202–15.
2. European Commission. European Guidelines for quality assurance in colorectal cancer screening and diagnosis. 2010.
3. Dobrow MJ, Hagens V, Chafe R, Sullivan T, Rabeneck L. Consolidated principles for screening based on a systematic review and consensus process. CMAJ. Canadian Medical Association; 2018; 190(14):E422–9.
4. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. Lancet. Elsevier; 2019; 394(10207):1467–80.
5. Schreuders EH, Ruco A, Rabeneck L, Schoen RE, Sung JJY, Young GP, et al. Colorectal cancer screening: a global overview of existing programmes. Gut. BMJ Publishing Group; 2015; 64(10):1637–49.
6. Lauby-Secretan B, Vilahur N, Straif K. IARC Handbook of Cancer Prevention Vol. 17 - Colorectal Cancer Screening. Journal of Global Oncology. 2018.
7. Nielsen KT. DCCG national guidelines on CRC screening [Danish: DCCG's nationale retningslinier for diagnostik og behandling af kolorektal cancer - Screening] [Internet]. 2014. Available from: http://www.dccg.dk/retningslinjer/august2014/2014_screening.pdf.
8. Klerk C de, Gupta S, Dekker E, Essink-Bot M, Expert Working Group 'Coalition to reduce inequities in colorectal cancer screening' of the World Endoscopy Organization. Socioeconomic and ethnic inequities within organised colorectal cancer screening programmes worldwide. Gut. 2017; gutjnl-2016-313311.
9. Jørgensen MD, Mikkelsen EM, Erichsen R, Thomsen MK. Mental illness and participation in colorectal cancer screening: a scoping review. Scand J Gastroenterol. Taylor & Francis; 2022.
10. Rasmussen M, Andersen ABT, Njor SH, Andersen VD. Danish Colorectal Cancer Screening Database, Annual report 2017 National Prevalence Screening Round [Danish] [Internet]. København/Århus; 2018 [cited 2019 Jan 15]. Available from: https://www.sundhed.dk/content/cms/45/61245_dts_årsrapport-2017_final.pdf.
11. Mikkelsen EM, Thomsen MK, Tybjerg J, Friis-Hansen L, Andersen B, Jørgensen JCR, et al. Colonoscopy-related complications in a nationwide immunochemical fecal occult blood test-based colorectal cancer screening program. Clin Epidemiol. Dove Press; 2018; Volume 10:1649–55.
12. Lawler M, Alsina D, Adams RA, Anderson AS, Brown G, Fearnhead NS, et al. Critical research gaps and recommendations to inform research prioritisation for more effective prevention and improved outcomes in colorectal cancer. Gut. 2018; 671136:179–93.
13. Schreuders EH, Grobbee EJ, Spaander MCW, Kuipers EJ. Advances in Fecal Tests for Colorectal Cancer Screening. Curr Treat Options Gastroenterol. Springer; 2016; 14(1):152–62.
14. Fraser CG. Faecal haemoglobin concentration and personalised assessment of the risk of colorectal neoplasia. J Lab Precis Med. 2017; 2:71–71.

15. Cooper JA, Moss SM, Smith S, Seaman HE, Taylor-Phillips S, Parsons N, et al. FIT for the future: a case for risk-based colorectal cancer screening using the faecal immunochemical test. Color Dis. Wiley/Blackwell (10.1111); 2016; 18(7):650–3.

16. Fitzpatrick-Lewis D, Ali MU, Warren R, Kenny M, Sherifali D, Raina P. Screening for colorectal cancer: A systematic review and meta-analysis. Clin Colorectal Cancer. Elsevier Inc.; 2016.

17. Hol L, Leerdam ME Van, Ballegooijen M Van, Vuuren AJ Van, Dekken H Van, Reijerink JCIY, et al. Screening for colorectal cancer: randomised trial comparing guaiac-based and immunochemical faecal occult blood testing and flexible sigmoidoscopy. Gut. Gut; 2010; 59(1):62–8.

18. Guittet L, Bouvier V, Mariotte N, Vallee JP, Arsène D, Boutreux S, et al. Comparison of a guaiac based and an immunochemical faecal occult blood test in screening for colorectal cancer in a general average risk population. Gut. Gut; 2007; 56(2):210–4.

19. Meklin J, Syrjänen K, Eskelinen M. Colorectal Cancer Screening With Traditional and New-generation Fecal Immunochemical Tests: A Critical Review of Fecal Occult Blood Tests. Anticancer Res. International Institute of Anticancer Research; 2020; 40(2):575–81.

20. Santos Silva I Dos. Cancer Prevention. In: Cancer Epidemiology: Principles and Methods. [Internet]. Lyon.; 1999. Available from: https://publications.iarc.fr/Non-Series-Publications/Other-Non-Series-Publications/Cancer-Epidemiology-Principles-And-Methods-1999.

21. Loong T. Understanding sensitivity and specificity with the right side of the brain. Bmj. 2003; 327(September):716–9.

22. Fletcher RH, Fletcher SW, Fletcher GS. Prevention. Methodologic issues in evaluating screening programs. In: Clinical Epidemiology the essentials. 5th ed. Wolters Kluwer Health; 2014; p. 159–62.

23. Young A, Hobbs R, Kerr D, editors. ABC of Colorectal Cancer. 2nd ed. Wiley-Blackwell; 2011.

24. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. International Agency for Research on Cancer [Internet]. Lyon, France; 2013 [cited 2016 Feb 9]. Available from: http://globocan.iarc.fr.

25. Nordcan. Cancer stat fact sheets Denmark - Colorectal. 09.07.2015 [Internet]. 2015 [cited 2015 Aug 27]. Available from: http://www-dep.iarc.fr/NORDCAN/english/StatsFact.asp?cancer=550&country=208.

26. Terry MB, Neugut AI, Bostick RM, Sandler RS, Haile RW, Jacobson JS, et al. Risk Factors for Advanced Colorectal Adenomas - A Pooled Analysis. Cancer Epidemiol. Biomarkers Prev. American Association for Cancer Research; 2002; 6(9):671–5.

27. International Agency for Research on Cancer (IARC). Rates per 100 000, incidence, males and females, age [40-85+], in 2013 Colorectum, Denmark [Internet]. [cited 2022 Sep 17]. Available from: https://gco.iarc.fr/overtime/en/dataviz/.

28. Cancer Research UK. Bowel Cancer Survival By Stage At Diagnosis. Bowel cancer survival statistics [Internet]. 2014 [cited 2015 Aug 25]. Available from:

http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/survival#heading-Three.

29. Stower MJ, Hardcastle JD. The results of 1115 patients with colorectal cancer treated over an 8-year period in a single hospital. Eur J Surg Oncol. 1985; 11(2):119–23.

30. NORDCAN. Survival: tables by cancer sites [Danish] [Internet]. [cited 2022 Sep 17]. Available from: https://www-dep.iarc.fr/nordcan/dk/frame.asp.

31. Zheng W, Rutter CM. Estimated mean sojourn time associated with hemoccult SENSA for detection of proximal and distal colorectal cancer. Cancer Epidemiol Biomarkers Prev. 2012; 21(10):1722–30.

32. Zhou L, Yu H, Zheng S. Evaluation of a fecal occult blood test with reverse passive hemagglutination for colorectal neoplasm screening. World J Gastroenterol. Baishideng Publishing Group Inc; 1997; 3(1):38.

33. Mertz-Petersen M, Piper TB, Kleif J, Ferm L, Christensen IJ, Nielsen HJ, et al. Triage for selection to colonoscopy? Eur J Surg Oncol. 2018; 44(10):1539–41.

34. Perri F, Iacobellis A, Gentile M, Tumino E, Andriulli A. The intelligent, painless, " germ-free" colonoscopy: A Columbus' egg for increasing population adherence to colorectal cancer screening? Dig Liver Dis. 2010; 42:839–43.

35. Slater A. Imaging of colorectal cancer. In: Young A, Hobbs R, Kerr D, editors. ABC of colorectal cancer. 2nd ed. BMJ books - Wiley-Blackwell; 2011; p. 31–5.

36. Lund M, Erichsen R, Valori R, Møller Jensen T, Helle Njor S, Laurberg S, et al. Data quality and colonoscopy performance indicators in the prevalent round of a FIT-based colorectal cancer screening program. Scand J Gastroenterol. Taylor and Francis Ltd; 2019; 54(4):471–7.

37. Rasmussen M, Ragner AZK, Njor SH, Andersen VD. Danish Colorectal Cancer Screening Database, Annual report 2020 [Danish] [Internet]. 2021 [cited 2022 Aug 21]. Available from: https://www.sundhed.dk/sundhedsfaglig/kvalitet/kliniske-kvalitetsdatabaser/screening/dansk-tarmkraeftscreeningsdatabase/.

38. Thomsen MK, Njor SH, Rasmussen M, Linnemann D, Andersen B, Baatrup G, et al. Validity of data in the Danish colorectal cancer screening database. Clin Epidemiol. 2017; 9.

39. Schmidt M, Schmidt SAJ, Adelborg K, Sundbøll J, Laugesen K, Ehrenstein V, et al. The Danish health care system and epidemiological research: from health care contacts to database records. Clin Epidemiol. Dove Medical Press Ltd; 2019; Volume 11:563–91.

40. Deding U, Henig AS, Salling A, Torp-Pedersen C, Bøggild H. Sociodemographic predictors of participation in colorectal cancer screening. Int J Colorectal Dis. Springer Berlin Heidelberg; 2017; 32(8):1117–24.

41. Larsen MB, Mikkelsen E, Rasmussen M, Friis-Hansen L, Ovesen A, Rahr HB, et al. Sociodemographic characteristics of nonparticipants in the Danish colorectal cancer screening program: a nationwide cross-sectional study. Clin Epidemiol. Dove Press; 2017; Volume 9:345–54.

42. Deding U, Vardinghus-Nielsen H, Torp-Pedersen C, Bøggild H. Health educational background as a predictor of non-participation in national colorectal cancer screening: A cross-sectional population study among 886,088 invited Danes. Prev Med (Baltim). 2019;

125.

43. Clarke RB, Therkildsen C, Gram MA, Andersen KK, Mørch LS, Tybjerg AJ. Use of primary health care and participation in colorectal cancer screening – a Danish national register-based study. Acta Oncol (Madr). Taylor & Francis; 2020; 1–5.

44. Marie Horshauge P, Gabel P, Bach Larsen M, Kirkegaard P, Edwards A, Andersen B. The association between health literacy and colorectal cancer screening uptake in a publicly funded screening program in Denmark: Cross-sectional study. Prev Med Reports. Elsevier BV; 2020; 19:101132.

45. Pallesen AVJ, Herrstedt J, Westendorp RGJ, Mortensen H, Kristiansen M. Differential effects of colorectal cancer screening across sociodemographic groups in Denmark: a register-based study. Acta Oncol ISSN. 2021.

46. Nielsen J, Berg-Beckhoff G, Leppin A. To do or not to do - a survey study on factors associated with participating in the Danish screening program for colorectal cancer. BMC Health Serv Res. BMC Health Serv Res; 2021; 21(1).

47. Deding U, Henig AS, Torp-Pedersen C, Bøggild H. The effects of reminders for colorectal cancer screening: participation and inequality. Int J Colorectal Dis. Springer Verlag; 2019; 34(1):141–50.

48. Deding U, Torp-Pedersen C, Bøggild H. The association between immigration status and ineligible stool samples for colorectal cancer screening. Cancer Epidemiol. Elsevier Ltd; 2018; 57:74–9.

49. Njor SH, Andersen B, Friis-Hansen L, Haas N de, Linnemann D, Nørgaard H, et al. The optimal cut-off value in fit-based colorectal cancer screening: An observational study. Cancer Med. Blackwell Publishing Ltd; 2021; 10(5).

50. Deding U, Henig AS, Hindersson P, Torp-Pedersen C, Bøggild H. Determinants of non-participation in colon examination following positive stool sample in colorectal cancer screening. Eur J Public Health. 2019; 29(6):1118–24.

51. Thomsen MK, Rasmussen M, Njor SH, Mikkelsen EM. Demographic and comorbidity predictors of adherence to diagnostic colonoscopy in the Danish Colorectal Cancer Screening Program: a nationwide cross-sectional study. Clin Epidemiol. Dove Press; 2018; Volume 10:1733–42.

52. Bie AKL, Brodersen J. Why do some participants in colorectal cancer screening choose not to undergo colonoscopy following a positive test result? A qualitative study. Scand J Prim Health Care. 2018; 36(3):262–71.

53. Larsen MB, Njor SH, Jensen TM, Ingeholm P, Andersen B. Potential for prevention: a cohort study of colonoscopies and removal of adenomas in a FIT-based colorectal cancer screening programme. Scand J Gastroenterol. 2019; 54(8):1008–14.

54. Larsen MB, Njor S, Ingeholm P, Andersen B. Effectiveness of Colorectal Cancer Screening in Detecting Earlier-Stage Disease-A Nationwide Cohort Study in Denmark. Gastroenterology. Gastroenterology; 2018; 155(1):99–106.

55. Njor SH, Larsen MB, Søborg B, Andersen B. Colorectal cancer mortality after randomized implementation of FIT-based screening - a nationwide cohort study. J Med Screen. SAGE Publications, Sage UK: London, England; 2022.

56. Cardoso R, Guo F, Heisser T, Hackl M, Ihle P, Schutter H De, et al. Colorectal cancer incidence, mortality, and stage distribution in European countries in the colorectal cancer screening era: an international population-based study. Lancet Oncol. Elsevier; 2021; 0(0).

57. Wilhelmsen M, Njor SH, Roikjær O, Rasmussen M, Gögenur I. Impact of screening on short-term mortality and morbidity following treatment for colorectal cancer. Scand J Surg. Scand J Surg; 2021.

58. Owen GS, Richardson G, David AS, Szmukler G, Hayward P, Hotopf M. Mental capacity to make decisions on treatment in people admitted to psychiatric hospitals: cross sectional study. BMJ [Internet]. British Medical Journal Publishing Group; 2008 [cited 2022 Sep 20]; 337. Available from: https://www.bmj.com/content/337/bmj.39580.546597.BE.

59. Howard LM, Barley EA, Davies E, Rigg A, Lempp H, Rose D, et al. Cancer diagnosis in people with severe mental illness: practical and ethical issues. Lancet Oncol. Elsevier; 2010; 11(8):797–804.

60. Solmi M, Firth J, Miola A, Fornaro M, Frison E, Fusar-Poli P, et al. Disparities in cancer screening in people with mental illness across the world versus the general population: prevalence and comparative meta-analysis including 4 717 839 people. The Lancet Psychiatry. Elsevier Ltd; 2020; 7(1):52–63.

61. Steele RJ, McDonald PJ, Digby J, Brownlee L, Strachan JA, Libby G, et al. Clinical outcomes using a faecal immunochemical test for haemoglobin as a first-line test in a national programme constrained by colonoscopy capacity. United Eur Gastroenterol J. SAGE Publications Ltd; 2013; 1(3):198–205.

62. Toes-Zoutendijk E, Leerdam ME van, Dekker E, Hees F van, Penning C, Nagtegaal I, et al. Real-Time Monitoring of Results During First Year of Dutch Colorectal Cancer Screening Program and Optimization by Altering Fecal Immunochemical Test Cut-Off Levels. Gastroenterology. 2017; 152(4):767-775.e2.

63. Terhaar sive Droste JS, Oort FA, Hulst RWM van der, Heukelem HA van, Loffeld RJLF, Turenhout ST van, et al. Higher fecal immunochemical test cutoff levels: lower positivity rates but still acceptable detection rates for early-stage colorectal cancers. Cancer Epidemiol Biomarkers Prev. American Association for Cancer Research; 2011; 20(2):272–80.

64. Grobbee EJ, Wieten E, Hansen BE, Stoop EM, Wijkerslooth TR, Lansdorp-Vogelaar I, et al. Fecal immunochemical test-based colorectal cancer screening: The gender dilemma. United Eur Gastroenterol J. 2017; 5(3):448–54.

65. Arana-Arri E, Idigoras I, Uranga B, Pérez R, Irurzun A, Gutiérrez-Ibarluzea I, et al. Population-based colorectal cancer screening programmes using a faecal immunochemical test: Should faecal haemoglobin cut-offs differ by age and sex? BMC Cancer. 2017; 17(1).

66. Alvarez-Urturi C, Andreu M, Hernandez C, Perez-Riquelme F, Carballo F, Ono A, et al. Impact of age- and gender-specific cut-off values for the fecal immunochemical test for hemoglobin in colorectal cancer screening. Dig Liver Dis. W.B. Saunders; 2016; 48(5):542–51.

67. Sarkeala T, Färkkilä M, Anttila A, Hyöty M, Kairaluoma M, Rautio T, et al. Piloting gender-oriented colorectal cancer screening with a faecal immunochemical test: population-based registry study from Finland. BMJ Open. British Medical Journal Publishing Group; 2021; 11(2):e046667.

68. Digby J, Fraser CG, Carey FA, McDonald PJ, Strachan JA, Diament RH, et al. Faecal haemoglobin concentration is related to severity of colorectal neoplasia. J Clin Pathol. BMJ Publishing Group; 2013; 66(5):415–9.

69. Collins GS, Reitsma JB, Altman DG, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med. BioMed Central; 2015; 13(1):1.

70. Stegeman I, Wijkerslooth TR De, Stoop EM, Leerdam ME Van, Dekker E, Ballegooijen M Van, et al. Combining risk factors with faecal immunochemical test outcome for selecting CRC screenees for colonoscopy. Gut. 2014; 63(3):466–71.

71. Cooper JA, Parsons N, Stinton C, Mathews C, Smith S, Halloran SP, et al. Risk-adjusted colorectal cancer screening using the FIT and routine screening data: Development of a risk prediction model. Br J Cancer. Nature Publishing Group; 2018; 118(2):285–93.

72. Yen AMF, Chen SLS, Chiu SYH, Fann JCY, Wang PE, Lin SC, et al. A new insight into fecal hemoglobin concentration-dependent predictor for colorectal neoplasia. Int J Cancer. 2014; 135(5):1203–12.

73. Stegeman I, Wijkerslooth TR de, Stoop EM, Leerdam ME van, Dekker E, Ballegooijen M van, et al. Combining risk factors with faecal immunochemical test outcome for selecting CRC screenees for colonoscopy. Gut. BMJ Publishing Group; 2014; 63(3):466–71.

74. Tinmouth J, Vella ET, Baxter NN, Dubé C, Gould M, Hey A, et al. Colorectal Cancer Screening in Average Risk Populations: Evidence Summary. Can J Gastroenterol Hepatol. Hindawi Limited; 2016; 2016.

75. Zorzi M, Fedeli U, Schievano E, Bovo E, Guzzinati S, Baracco S, et al. Impact on colorectal cancer mortality of screening programmes based on the faecal immunochemical test. Gut. 2015; 64(5).

76. Wiegering A, Ackermann S, Riegel J, Dietz UA, Götze O, Germer C-T, et al. Improved survival of patients with colon cancer detected by screening colonoscopy. Int J Colorectal Dis. 2016; 31(5):1039–45.

77. Spolverato G, Capelli G, Battagello J, Barina A, Nordio S, Finotti E, et al. More Favorable Short and Long-Term Outcomes for Screen-Detected Colorectal Cancer Patients. Front Oncol. 2021; 11:620644.

78. Losurdo P, Giacca M, Biloslavo A, Fracon S, Sereni E, Giudici F, et al. Colorectal cancer-screening program improves both short- and long-term outcomes: a single-center experience in Trieste. Updates Surg. 2020; 72(1):89–96.

79. Mengual-Ballester M, Pellicer-Franco E, Valero-Navarro G, Soria-Aledo V, García-Marín JA, Aguayo-Albasini JL. Increased survival and decreased recurrence in colorectal cancer patients diagnosed in a screening programme. Cancer Epidemiol. 2016; 43:70–5.

80. Paszat LF, Sutradhar R, Corn E, Luo J, Baxter NN, Tinmouth J, et al. Morbidity and mortality following major large bowel resection for colorectal cancer detected by a population-based screening program. J Med Screen. SAGE Publications, Sage UK: London, England; 2020.

81. Njor SH, Friis-Hansen L, Andersen B, Søndergaard B, Linnemann D, Jørgensen JCR, et al. Three years of colorectal cancer screening in Denmark. Cancer Epidemiol. Elsevier Ltd;

2018; 57:39–44.

82. Brenner H, Chang-Claude J, Seiler CM, Rickert A, Hoffmeister M. Protection from colorectal cancer after colonoscopy: A population-based, case-control study. Ann Intern Med. American College of Physicians; 2011; 154(1):22–30.

83. Ingeholm P, Gögenür I, Iversen L. Danish Colorectal Cancer Group Database. Clin Epidemiol. Dove Press; 2016; Volume 8:465–8.

84. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: A review of content, data quality, and research potential. Clin Epidemiol. 2015; 7:449–90.

85. Kildemoes HW, Sorensen HT, Hallas J. The Danish National Prescription Registry. Scand J Public Heal. 2011; 39(7 Suppl):38–41.

86. Helweg-Larsen K. The Danish Register of Causes of Death. Scand J Public Health. SAGE PublicationsSage UK: London, England; 2011; 39(7):26–9.

87. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc. Lippincott Williams and Wilkins; 2015; 13(3):141–6.

88. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018; 169(7):467–73.

89. Zeng X, Zhang Y, Kwong JSW, Zhang C, Li S, Sun F, et al. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. J Evid Based Med. John Wiley & Sons, Ltd; 2015; 8(1):2–10.

90. Larsen PT, Jørgensen SF, Njor SH. Colorectal cancer screening participation among citizens not recommended to be screened: a cohort study. BMC Gastroenterol 2022 221. BioMed Central; 2022; 22(1):1–7.

91. Gram MA, Therkildsen C, Clarke RB, Andersen KK, Mørch LS, Tybjerg AJ. The influence of marital status and partner concordance on participation in colorectal cancer screening. Eur J Public Health. Oxford University Press (OUP); 2020; 1–6.

92. Plana-Ripoll O, Pedersen CB, Agerbo E, Holtz Y, Erlangsen A, Canudas-Romo V, et al. A comprehensive analysis of mortality-related health metrics associated with mental disorders: a nationwide, register-based cohort study. Lancet. 2019; 394(10211):1827–35.

93. Momen NC, Plana-Ripoll O, Agerbo E, Christensen MK, Iburg KM, Laursen TM, et al. Mortality Associated With Mental Disorders and Comorbid General Medical Conditions. JAMA Psychiatry. American Medical Association; 2022; 79(5):444–53.

94. Bambra C. Health inequalities and welfare state regimes: theoretical insights on a public health "puzzle." J Epidemiol Community Heal. 2011; 65(9):740–5.

95. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ. British Medical Journal Publishing Group; 2009; 338:b375.

96. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research:

Developing a prognostic model. BMJ. British Medical Journal Publishing Group; 2009; 338:b604.

97. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. BMJ. British Medical Journal Publishing Group; 2009; 338(may28 1):b605–b605.

98. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. Oxford University Press; 2014; 35(29):1925–31.

99. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology. 2010; 21(1):128–38.

100. Moscoe E, Bor J, Bärnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. J Clin Epidemiol. 2015; 68(2):132–43.

101. Parner ET, Andersen PK. Regression analysis of censored data using pseudo-observations. Stata J. 2010; 10(3):408–22.

102. Collins GS, Groot JA de, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol. BioMed Central; 2014; 14(1):40.

103. Danish Health Authority. Recommendations regarding screening for colorectal cancer [Danish] [Internet]. 2010. Available from: https://www.sst.dk/-/media/Udgivelser/2012/Publ2012/Anbefalninger-vedrørende-screening-for-tyk--og-endetarmskræft.ashx.

104. Bor J, Moscoe E, Mutevedzi P, Newell ML, Bärnighausen T. Regression discontinuity designs in epidemiology: Causal inference without randomized trials. Epidemiology. Lippincott Williams and Wilkins; 2014; 25(5):729–37.

105. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology. 2010; 21(3):383–8.

106. Thomsen MK, Pedersen L, Erichsen R, Lash TL, Sørensen HT, Mikkelsen EM. Risk-stratified selection to colonoscopy in FIT colorectal cancer screening: development and temporal validation of a prediction model. Br J Cancer 2022. Nature Publishing Group; 2022; 1–7.

107. Cairns JM, Greenley S, Bamidele O, Weller D. A scoping review of risk-stratified bowel screening: current evidence, future directions. Cancer Causes Control. Springer; 2022; 33(5):653–85.

108. Kortlever TL, Vlugt M van der, Dekker E, Bossuyt PMM. Individualized faecal immunochemical test cut-off based on age and sex in colorectal cancer screening. Prev Med Reports. 2021; 23:101447.

109. Roos VH, Kallenberg FGJ, Vlugt M van der, Bongers EJC, Aalfs CM, Bossuyt PMM, et al. Addition of an online, validated family history questionnaire to the Dutch FIT-based screening programme did not improve its diagnostic yield. Br J Cancer. Nature Publishing Group; 2020; 122(12):1865–71.

110. Cooper JA, Ryan R, Parsons N, Stinton C, Marshall T, Taylor-Phillips S. The use of electronic healthcare records for colorectal cancer screening referral decisions and risk prediction model development. BMC Gastroenterol. BioMed Central; 2020; 20(1):1–16.

111. Duran-Sanchon S, Moreno L, Gómez-Matas J, Augé JM, Serra-Burriel M, Cuatrecasas M, et al. Fecal MicroRNA-Based Algorithm Increases Effectiveness of Fecal Immunochemical Test–Based Screening for Colorectal Cancer. Clin Gastroenterol Hepatol. W.B. Saunders; 2021; 19(2):323-330.e1.

112. Kadiyala S, Strumpf E. How Effective is Population-Based Cancer Screening? Regression Discontinuity Estimates from the US Guideline Screening Initiation Ages. Forum Heal Econ Policy. Walter de Gruyter GmbH; 2016; 19(1):87–139.

113. Kadiyala S, Strumpf EC. Are United States and Canadian cancer screening rates consistent with guideline information regarding the age of screening initiation? Int J Qual Heal Care. 2011; 23(6):611–20.

114. Sashegyi A, Ferry D. On the Interpretation of the Hazard Ratio and Communication of Survival Benefit. Oncologist. Oxford Academic; 2017; 22(4):484–6.

115. Lash TL, Fox MP, Fink AK. Introduction, Objectives and an Alternative. In: Applying quantitative bias analysis to epidemiologic data. 1st ed. Springer; 2010; p. 1–12.

116. Domino ME, Beadles CA, Lichstein JC, Farley JF, Morrissey JP, Ellis AR, et al. Heterogeneity in the quality of care for patients with multiple chronic conditions by psychiatric comorbidity. Med Care. 2014; 52.

117. Druss BG, Rask K, Katon WJ. Major Depression, Depression Treatment, and Quality of Primary Medical Care. Gen Hosp Psychiatry. NIH Public Access; 2008; 30(1):20.

118. Druss BG, Rosenheck RA, Desai MM, Perlin JB. Quality of preventive medical care for patients with mental disorders. Med Care. Med Care; 2002; 40(2):129–36.

119. Folsom DP, McCahill M, Bartels SJ, Lindamer LA, Ganiats TG, Jeste D V. Medical comorbidity and receipt of medical care by older homeless people with schizophrenia or depression. Psychiatr Serv. American Psychiatric Publishing; 2002; 53(11):1456–60.

120. Holden DJ, Jonas DE, Porterfield DS, Reuland D, Harris R. Systematic Review: Enhancing the Use and Quality of Colorectal Cancer Screening. Ann Intern Med. 2010; 152(10):668.

121. Stecker T, Fortney JC, Prajapati S. How Depression Influences The Receipt of Primary Care Services among Women: A Propensity Score Analysis. J women's Heal. Mary Ann Liebert, Inc. 2 Madison Avenue Larchmont, NY 10538 USA; 2007; 16(2):198–205.

122. Rothman KJ, Greenland S, Lash TL. Validity in Epidemiologic Studies. In: Modern Epidemiology. 3rd ed. Lippincott Williams & Wilkins; 2012.

123. Lash TL, Fox MP, Fink AK. Selection bias. In: Applying quantitative bias analysis to epidemiologic data. 1st ed. Springer; 2010.

124. Olsen J. Using Secondary Data. In: Rothman KJ, Greenland S, Lash TL, editors. Modern Epidemiology. 3rd ed. Philadelphia USA: Lippincott Williams & Wilkins; 2012; p. 481–91.

125. Ranganathan P, Pramesh C. Censoring in survival analysis: Potential for bias. Perspect Clin Res. 2012; 3(1):40.

126. Lash TL, Fox MP, Fink AK. Misclassification. In: Applying quantitative bias analysis to

epidemiologic data. 1st ed. Springer; 2010; p. 79–108.

127. Mors O, Perto GP, Mortensen PB. The Danish Psychiatric Central Research Register. Scand J Public Heal. 2011; 39:54–7.

128. Fletcher RH, Fletcher SW, Fletcher GS. Risk: Exposure to disease. In: Clinical Epidemiology the essentials. 5th ed. Wolters Kluwer Health; 2014; p. 61–79.

129. Spix C, Berthold F, Hero B, Michaelis J, Schilling FH. Correction factors for self-selection when evaluating screening programmes. J Med Screen. SAGE Publications Ltd; 2016; 23(1):44–9.

130. Bjerregaard JK, Fredsted A, Njor SH, Andersen VD. Danish Colorectal Cancer Group Database, Annual report 2020 [Danish] [Internet]. 2021 [cited 2022 Aug 21]. Available from: https://www.sundhed.dk/sundhedsfaglig/kvalitet/kliniske-kvalitetsdatabaser/kraeft/kolorektal-cancer-database/.

131. Yamada Y, Fujiwara M, Shimazu T, Etoh T, Kodama M, So R, et al. Patients' acceptability and implementation outcomes of a case management approach to encourage participation in colorectal cancer screening for people with schizophrenia: a qualitative secondary analysis of a mixed-method randomised clinical trial. BMJ Open. 2022; 12:60621.

132. Lansdorp-Vogelaar I, Meester R, Jonge L de, Buron A, Haug U, Senore C. Risk-stratified strategies in population screening for colorectal cancer. Int J Cancer. John Wiley & Sons, Ltd; 2022; 150(3):397–405.

133. Clark GRC, Steele RJC, Fraser CG. Strategies to minimise the current disadvantages experienced by women in faecal immunochemical test-based colorectal cancer screening. Clin Chem Lab Med. De Gruyter Open Ltd; 2022; 60(10):1496–505.

134. Christiansen SR, Autier P, Støvring H. Change in effectiveness of mammography screening with decreasing breast cancer mortality: a population-based study. Eur J Public Health. Oxford Academic; 2022; 32(4):630–5.

Appendix Table 1. Study characteristics and results of studies identified in literature search for Study III

Adapted from master's thesis.

Search string: Pubmed (February 27, 2020): "colorectal neoplasms"[MeSH] AND screening[Title/Abstract] AND ("hemoglobin"[Title/Abstract] OR "haemoglobin"[tiab] OR "FIT"[tiab] OR "fecal immunochemial test"[tiab] or) AND (predict*[Title/Abstract]) AND ("humans"[MeSH] AND English[lang])

Author, *title.* Year [ref] country. FIT brand.	Design and setting, predictors	Results	Limitations/comments
Stegeman et al., *Combining risk factors with faecal immunochemical test outcome for selecting CRC screenees for colonoscopy*. 2013 (70) The Netherlands. FIT: OC-sensor (Eiken).	Study population: 1,112 participants in colonoscopy-based screening whom also delivered a sample for FIT analyses. Risk factor information by self-reported questionnaire. Predictors: Total calcium intake, family history of CRC, age and fHb. Outcome: AN.	Calibration slope was 1 and AUC for discrimination was 76%. They did a Net Reclassification Improvement (NRI) analysis, which sums reclassifications in the right direction by two methods. Compared to a FIT cut-off at 10 µg Hb/g feces the NRI was 0.054 in favor of the prediction model.	Prediction model was not validated in separate data. Small sample size.
Yen et al., *A new insight into fecal hemoglobin concentration-dependent predictor for colorectal neoplasia*. 2014 (72) Taiwan. FIT: OC-sensor (Eiken).	Study population: 54,921 participants in a program offering annual FIT screening to ≥40 years (cut-off 20 µg Hb/g feces, but all were included in study). Time to event analysis. External validation using data from two	AUC of the models with FIT only was 83.0% (95% CI: 81.5; 84.4) for AN and 84.3% (95% CI: 81.9;86.7) for CRC. The model predicting CRC had lower AUC in external validation (AUC= 78.7% (95% CI:77.0;80.4)), whereas the model predicting	Description of methods unclear (e.g. follow-up start and end, definition of predictors and variable selection). Outcome was AN both after a negative or positive FIT, which occurred after a mean of 4.25 years, thus a mix of disease detected at baseline screening, at subsequent screening

	other community programs. Model 1: risk factors (Gender, Family history of CRC, diabetes, hyper-tension, alcohol, smoking, BMI, TG, TC) Model 2: fHb. Model 3: model 1+2. Outcome: AN.	AN performed similarly in the validation dataset. AUC increased only slightly when including the conventional risk factors.	rounds, interval cancers or clinically detected possibly years after last screening – risk of differential misclassification. Do not evaluate calibration.
Cooper et al., *Risk-adjusted colorectal cancer screening using the FIT and routine screening data:* *development of a risk prediction model*. 2018 (71) United Kingdom. FIT: OC-sensor (Eiken).	Study population: 1,810 individuals 59-74 years with a positive FIT (≥20 µg Hb/g feces) who underwent colonoscopy. Cross-validation for internal validity. Predictors: Age, gender, participation in last screening round and fHb. Outcome: AN.	Their models had AUCs of 0.63 (95% CI: 0.60–0.66) for FIT only and 0.66 (95% CI: 0.63–0.69) for the risk-adjusted model. Calibration was good for both models. At all corresponding FIT- and risk cut-offs, sensitivity was higher for the risk method.	They calculate sensitivity, even though they only have data on those with a FIT ≥20 µg Hb/g feces; in this case the number of false negatives in individuals with fHb below 20 µg Hb/g feces is not known. No external validation.

Abbreviations: fHb=fecal hemoglobin, AN=advanced neoplasia, AA=advanced adenoma, OR=odds ratio, CRC=colorectal cancer, AUC=area under the curve.

The papers have been removed from the file due to copyright issues

Paper I

Mental illness and participation in colorectal cancer screening:

a scoping review

Published in Scandinavian Journal of Gastroenterology, May 2022

Paper II

Mental disorders and colorectal cancer screening participation and trajectories – a Danish cohort study

Paper III

Risk-stratified selection to colonoscopy in FIT colorectal cancer screening: development and temporal validation of a prediction model

Published in British Journal of Cancer, January 2022

Paper IV

Effect of a FIT-based Colorectal Cancer screening program on mortality estimated by the Regression Discontinuity Design