

# Evaluating benefits and harms of screening

- the streetlight effect?

Minna Johansson

Department of Public Health  
and Community Medicine  
/Primary Health Care,  
Institute of Medicine at  
Sahlgrenska Academy  
University of Gothenburg

Gothenburg, Sweden, 2018



UNIVERSITY OF  
GOTHENBURG

Cover illustration by Sebastian Dijkstra

*Evaluating benefits and harms of screening – the streetlight effect?*

© 2018 Minna Johansson

[minna.johansson@vgregion.se](mailto:minna.johansson@vgregion.se)

ISBN 978-91-7833-003-4

Printed in Gothenburg, Sweden 2018

BrandFactory AB

To my grandfather Sigvard  
for the joy in your eyes when you looked at me  
and for the courage you gave me



# Abstract

The general aim of this thesis was to explore how the benefits and harms of screening for a potentially life-threatening disease can be evaluated.

Papers I and II are a Cochrane Systematic Review on screening for malignant melanoma. We found no randomised trials of the benefits and harms of screening for malignant melanoma. We concluded that due to the uncertainty of benefits, and risk of harms through overdiagnosis and opportunity costs, screening for malignant melanoma should not be recommended outside the confines of a well-designed, randomised trial. However, screening for malignant melanoma is already widely adopted in the Western world.

Papers III and IV explore screening for abdominal aortic aneurysm (AAA). In study III, we found that AAA screening has been introduced in several countries without adequate investigation of harms. We also found that AAA screening caused harm through the detection and subsequent surgery of AAAs that would never have caused symptoms (i.e. overdiagnosis and overtreatment). Study IV is a registry study of the benefits and harms of AAA screening in Sweden. We found that AAA-mortality in Swedish men aged 65-74 has dropped by about 70% in the last decades. Screening had, at best, a minor effect on the decline in AAA-mortality, which was likely caused mainly by reduced smoking. We estimated that for every 10 000 men invited, 2 men (95% CI -3 to 7) avoided AAA-death (not statistically significant). At the same time, 49 men were likely overdiagnosed (95% CI 25 to 73), of whom 19 men (95% CI 1 to 37) had unnecessary surgery with a risk of mortality and morbidity. The remaining 30 men were offered regular follow-up with potential psychosocial consequences. The effect on AAA-mortality in Sweden was only 7% of that in the largest randomised trial. The less favourable benefit-to-harm balance brings into question the continued use of AAA screening.

The overall conclusion of this thesis is that benefits of screening receive much more attention and appreciation than harms.

Key Words: Abdominal aortic aneurysm, Benefits, Harms, Malignant melanoma, Mortality, Overdiagnosis, Overtreatment, Screening

# Sammanfattning på svenska

Poängen med screening är att hitta och behandla farlig sjukdom i ett tidigt stadium. För de flesta av oss låter detta intuitivt tilltalande. Men när vi undersöker människor utan symtom riskerar vi att orsaka skada. Det beror framförallt på att många av oss lever med avvikelser i våra kroppar som skulle tolkas som sjukdom om de hittades, men som aldrig kommer att ge några symtom. Problemet är att vi ofta inte kan skilja de avvikelser som kommer att orsaka sjukdom från de som aldrig kommer att orsaka sjukdom. Därför leder screening till att vissa människor i onödan får en diagnos (överdiagnostik) och att vissa av dem i onödan genomgår en behandling (överbehandling).

I denna avhandling har för- och nackdelar med screening för malignt melanom och screening för bukaortaaneurysm undersökts.

Malignt melanom är en cancer i huden som kan orsaka död genom metastaser till andra organ. Screening för melanom görs oftast genom att en läkare tittar över huden hos människor utan misstanke om melanom. Screening för melanom är utbredd i västvärlden. Vi undersökte om det finns vetenskapliga belägg från randomiserade studier för att screening för melanom leder till mer nytta än skada, vilket normalt är ett krav för att screening ska rekommenderas. Vi fann att det saknas belägg för att screening leder till någon nytta i form av minskad död i melanom. Det finns vissa belägg från befolkningsstatistik för att screening leder till skada i form av överdiagnostik och överbehandling. Slutsatsen blev att screening för melanom inte bör rekommenderas.

Bukaortaaneurysm är en utvidgning av den stora kroppspulsådern. Buk-ortaaneurysm ger oftast inga symtom om det inte spricker. Om det spricker är dödligheten hög, över 80 procent. Screening för bukaortaaneurysm har införts i Sverige, Storbritannien och USA. Screeningen görs genom en ultraljuds-undersökning av magen hos 65-åriga män. Vi undersökte först om nackdelar med screening för bukaortaaneurysm har utvärderats och kom fram till att det saknas forskning på nackdelar med screeningen. Vi fann också att

screening orsakar skada genom att man hittar bukaortaaneurysm som aldrig skulle ha orsakat symtom om mannen i fråga inte hade deltagit i screening (överdiagnostik). Konsekvensen är att dessa män i onödan får leva med vetskapen om att de har ett bukaortaaneurysm, vilket kan påverka livskvalitén. Dessutom genomgår vissa av dessa män i onödan en operation för sitt aneurysm (överbehandling). Operationen har en risk för allvarliga komplikationer och död.

Screening för bukaortaaneurysm infördes efter att studier från 1980- och 90-talen visat att screening minskar risken att dö i bukaortaaneurysm. Sedan dess har förekomsten av bukaortaaneurysm minskat kraftigt. Detta beror troligen på minskad rökning. När förekomst och dödlighet av den sjukdom man screenar för minskar så minskar också nyttan av screening, och balansen mellan skada och nytta av screening försämras oftast. Vi undersökte därför nyttan (minskad död i bukaortaaneurysm) och skadan (överdiagnostik och överbehandling) av det svenska screeningprogrammet för bukaortaaneurysm. Vi fann att nyttan med screening har sjunkit kraftigt jämfört med de studier som låg till grund för att screeningen infördes. Balansen mellan skada och nytta har också försämrats eftersom de skadliga effekterna inte har minskat lika mycket. Vi beräknade att per 10 000 män som kallas till screening kommer 2 män undvika död i bukaortaaneurysm (men denna effekt var inte statistiskt säkerställd). Samtidigt kommer 49 män i onödan få en bukaortaaneurysm-diagnos (bli överdiagnosticerade) och 19 av dem kommer i onödan att genomgå en förebyggande operation med risk för död och allvarliga komplikationer (bli överbehandlade). Vi kom till slutsatsen att det är tveksamt om nyttan med screening för bukaortaaneurysm överväger skadan, och att det svenska screeningprogrammet därför bör omvärderas.

Malignt melanom och bukaortaaneurysm kan orsaka svår sjukdom och för tidig död. Det finns en stark tro bland forskare, politiker, vårdpersonal och allmänheten att screening, genom tidig upptäckt av dessa sjukdomar, ska kunna lindra detta lidande. Men förhoppningen att screening ska leda till nytta omsätts inte alltid i praktiken, och screening leder också till skada. När vården aktivt kallar människor till en undersökning som de inte har bett om är det viktigt att vi är säkra på att den undersökningen leder till mer nytta än skada. Denna avhandling är ägnad den uppgiften.

# List of papers

This thesis is based on the following papers:

- I. Johansson M, Brodersen J, Gøtzsche P, Jørgensen KJ.  
Screening for reducing morbidity and mortality in malignant melanoma (Protocol).  
Cochrane Database of Systematic Reviews 2016, Issue 9. Art No.:CD012352.
- II. Johansson M, Brodersen J, Gøtzsche P, Jørgensen KJ.  
Screening for reducing morbidity and mortality in malignant melanoma (Systematic Review).  
(Submitted)
- III. Johansson M, Hansson A, Brodersen J.  
Estimating overdiagnosis in screening for Abdominal Aortic Aneurysm: could a change in smoking habits and lowered aortic diameter tip the balance of screening towards harm?  
BMJ 2015;350:h825.
- IV. Johansson M, Zahl PH, Siersma V, Jørgensen KJ, Marklund B, Brodersen J.  
Benefits and harms of screening men for abdominal aortic aneurysm in Sweden - comparing age-matched, contemporary screened and non-screened cohorts in a population with falling incidence.  
(In press, the Lancet)



# List of related papers

Additional publications that are relevant to the General Discussion of this thesis (attached in Appendix):

1. Johansson M, Jørgensen KJ, Getz L, Moynihan R. Informed choice in a time of too much medicine – no panacea for ethical difficulties. *BMJ* 2016; 353:i2230.
2. Johansson M, Jørgensen KJ, Brodersen J. Harms of screening for abdominal aortic aneurysm: is there more to life than a 0.46% disease-specific mortality reduction? *Lancet* 2016;387:308-10.
3. Johansson M, Harris RP. Thresholds in women with abdominal aortic aneurysm. *Lancet* 2017;389:2446-8.
4. Johansson M, Brodersen J. Informed choice in screening needs more than information. *Lancet* 2015;385:1597-9.

# Content

Abstract.....	5
Sammanfattning på svenska.....	6
List of papers.....	8
List of related papers.....	9
Content.....	10
Prologue.....	13
Introduction.....	15
Screening – historical context.....	15
Benefits and harms of screening - taxonomy.....	23
Estimating benefits and harms of screening.....	29
Beyond benefits and harms.....	41
Screening for Malignant Melanoma.....	44
Screening for Abdominal Aortic Aneurysm.....	47
Aims.....	51
Methods.....	53
Results.....	57
Discussion of the studies.....	61
General discussion.....	67
The fundamental appeal of screening.....	68
Societal interventions vs screening individuals.....	74
Medical visions and steady progress.....	76
Consequences of the evidence hierarchy.....	77
Screening – deeper and broader meaning.....	78
The ethical dilemma of screening.....	81
Informed choices – a solution?.....	82
Conclusion.....	87
Future perspective.....	89
Epilogue.....	93

Acknowledgement.....	97
References.....	103
Papers I-IV .....	
Appendix.....	



# Prologue

My interest in the topic of this thesis started as a reaction to the current tendency within both research and policy-making to focus on benefits and disregard harms of preventive medicine in general, and screening in particular. I have learned a great deal about the significance of sound methodology and I have gained insights about the shortcomings of much of the evidence base on which Western medicine rests. I do believe that these aspects are immensely important. But in the process of writing this thesis, I have become aware that there are problems in medical science today that are much more profound than that.

In my opinion, the presumed value neutrality of medical research within our scientific paradigm is hugely problematic. Because; there is no such thing as value neutral science. What we do in research, all the way from the research questions we pose, to the research methods we chose, to the interpretation of our results, is all value-laden. We need to recognise that, and not hide behind a mask of scientific neutrality.

I believe that we as researchers should devote more attention to the humanistic and existential aspects of our work, and incorporate philosophy and sociology in our projects, to a much larger extent than what we do today. In my opinion, the way forward is to be much more sensitive to the inherent values of our endeavour. Instead of aiming for value neutrality, we should explore and critically reflect on the inherent values of our research, and make them explicit. We should take a step back, and view our own ideas within the landscape of a wider history of ideas.



# Introduction

## Screening – historical context

### **The rise of preventive medicine**

During the last half a century, Western medicine has evolved from primarily focusing on people with symptoms of disease to an increased attention on preventing or finding disease in asymptomatic people (Sackett 2002). In this context, mass screening to detect risk factors and early stages of disease has gained huge popularity amongst policy makers, health care personnel, the media and lay people (Schwartz 2004, Wegwarth 2012, Chen 2013).

To prevent people from getting sick in the first place, or to find disease early – even before symptoms, sounds intuitively appealing to most of us and there are many examples in medicine where this is fully justified. But this strategy entails specific considerations; the ethical premises of preventive medicine are fundamentally different compared to when people seek the health care system due to symptoms (Sackett 2002). David Sackett, often referred to as one of the founders of evidence based medicine, reflected on this in a paper from 2002 titled “The arrogance of preventive medicine”:

”...the 2 disciplines [”curative” and preventive medicine] are absolutely and fundamentally different in their obligations and implied promises to the individuals whose lives they modify. When patients sought me out for help with their established, symptomatic diseases, I promised them

only to do my best and never guaranteed that my interventions would make them better.. ..But surely the fundamental promise we make when we actively solicit individuals and exhort them to accept preventive interventions must be that, on average, they will be the better for it. Accordingly, the *presumption* that justifies the *aggressive assertiveness* with which we go after the unsuspecting healthy must be based on the highest level of randomized evidence that our preventive manoeuvre will, in fact, do more good than harm.”

He continues:

“First, it [preventive medicine] is aggressively assertive, pursuing symptomless individuals and telling them what they must do to remain healthy.. ..Second, preventive medicine is presumptuous, confident that the interventions it espouses will, on average, do more good than harm to those who accept and adhere to them. Finally, preventive medicine is overbearing, attacking those who question the value of its recommendations.”

Preventive medicine has, in only half a century, fundamentally changed Western medicine (Fugelli 2006, Pétursson 2012). However, our understanding of the meaning of diseases and diagnoses might not have adjusted to the new premises.

### **A new understanding of disease progression**

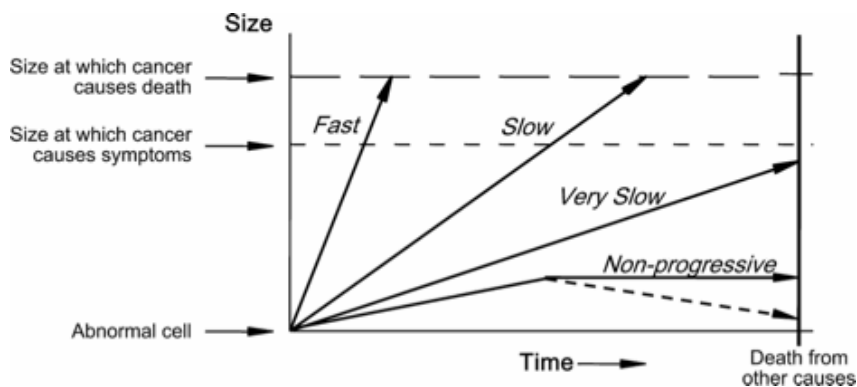
Historically, our understanding of disease progression has primarily been based on patients with symptoms of disease (Welch 2011). But today, people are diagnosed with life threatening diseases while being free of symptoms. This development has revealed that disease progression is widely diverse and utterly complex; our previous under-



standing of the natural history of many diseases has become challenged (Welch 2011).

### Non-predictable disease progression in cancer

Previously, cancer has been understood as a phenomenon that in most cases will lead to death if left untreated. But today, it is probably more accurate to understand a diagnosis of cancer as a pathological description made at a single time point, with diverse and sometimes limited relevance for the affected individual (Welch 2010). In the figure below from Welch and colleagues, a schematic presentation of the different ways in which cancer can develop is presented (Welch 2010).



**Figure 1.** Heterogeneity of cancer progression

Reproduced with permission from (Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst* 2010;102:605-13) Copyright © 2010, Oxford University Press.

Lethal cancer can grow fast or slow. Some cancers grow so slowly that they would never have caused symptoms in the remaining

lifespan, i.e. they are progressive, but not lethal. Additionally, some cancers are non-progressive and some cancers can even regress and disappear (Zahl 2008, Welch 2010). Paradoxically, screening tends to be better at finding the slow-growing or non-progressive cancers than the fast growing ones, because the time span for a detectable but asymptomatic stage is longer for these cancers (length bias) (Welch 2010).

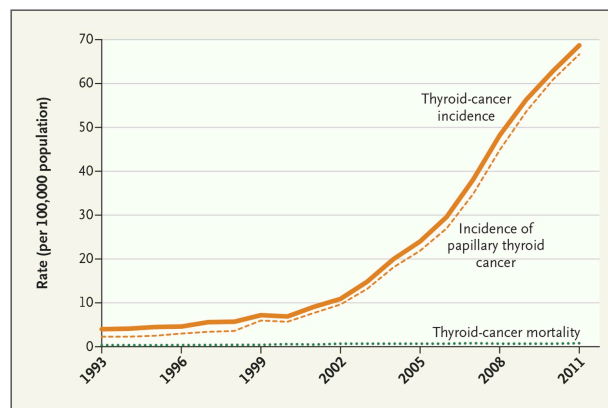
That cancer progression tends to be non-predictable presents a problem in a screening context (Zahl 2008, Welch 2010). This is because it is not possible to know if a cancer detected at screening would eventually cause problems or not. From autopsy studies, it is clear that for many cancers, there is an asymptomatic disease reservoir, which would have been labelled as cancer if it had been found, but nonetheless did not cause symptoms before death from other cause (Welch 2010). For some cancers, this is well-recognised and uncontroversial. For example, 30-70% of men above 60 years of age with no symptoms of prostate cancer, and who died from an unrelated cause, have been found to have histopathologically verified prostate cancer (Sakr 1996, Stamatiou 2006, Damiano 2007). Another autopsy study suggests that almost all adults would have a thyroid cancer if the thyroid gland would be thoroughly enough investigated (Harach 1985).

There are two reasons that an asymptomatic disease reservoir exists; either the cancer never progress (or even regress), or the cancer progress too slow to become symptomatic before the person dies from another cause (Zahl 2008, Welch 2010, Zahl 2013). It is worth noting that even highly aggressive cancers can remain asymptomatic throughout the lifespan of the affected person, if the person has high competing risks, i.e. a high risk of dying of something else in a short time-span (Zahl 2013).

The problems that non-predictable disease progression and disease reservoirs entail in a screening context can be exemplified with the cases of thyroid cancer and neuroblastoma.

#### Case study – thyroid cancer

Screening for thyroid cancer was increasingly adopted in South Korea from the late 1990s (Ahn 2014). As displayed in the graph below from Ahn and colleagues, thyroid-cancer incidence increased slowly during the 1990s, then rapidly in the 2000s. In 2011, the rate of thyroid-cancer diagnoses was 15 times the rate observed in 1993.



**Figure 2.** Thyroid-cancer incidence and related mortality in South Korea  
Reproduced with permission from (Ahn HS, Kim HJ, Welch HG. Korea's thyroid-cancer "epidemic" – screening and overdiagnosis. *N Engl J Med* 2014;371:1765-7) Copyright Massachusetts Medical Society.

Despite the dramatic increase in incidence, mortality from thyroid cancer has remained stable. In this time-span, there has been no im-

provement in the treatment of thyroid cancer. Additionally, effects from improved treatment seems unlikely to explain this pattern, since such effects would have to exactly and simultaneously equal out the increase in incidence throughout an extensive time period. Instead, it is suggestive of a large rate of detection of cancers that would not have caused symptoms if they would have remained undetected (Ahn 2014). In other words; screening has been finding cancers that did not need to be found. Similar patterns have been observed for kidney cancer, prostate cancer and malignant melanoma (Welch 2011).

#### Case study – neuroblastoma

Neuroblastoma is an extra-cranial solid cancer that affects children. Screening for neuroblastoma through measuring catecholamines in urine was introduced in Japan in the 1980s (Katanoda 2016). Later, non-randomised controlled trials revealed that mortality from neuroblastoma remained unchanged with screening (Schilling 2002, Woods 2002). However, screening resulted in a marked increase in the incidence of the disease. As neuroblastomas are extremely rare beyond adolescence, even comparatively short follow-up strongly suggested that screening found neuroblastomas that would later have disappeared by themselves if left undetected (Schilling 2002). Moreover, the earlier detection due to screening did not seem to have a beneficial effect on those neuroblastomas that would eventually have led to symptoms and death (Katanoda 2016).

Are the benefits from early detection of cancer exaggerated?

The prognosis of cancer is often closely linked to the stage of the tumour at diagnosis, which presents a rationale to believe that early detection of cancer is crucial. However, some evidence suggests that the

stage at diagnosis might be a consequence of the inherent biological aggressiveness of the tumour and not primarily a factor determined by the time of detection (Zahl 2008). Metastases might be formed very early in the tumours' development process, for those tumours with a biological predisposition to do so, i.e. even before they are detectable at screening (Zahl 2008). If this is true, the current emphasis of early detection of cancer might be exaggerated.

#### Non-predictable disease progression in non-cancerous disease

Much of the research in this area is based on cancer and cancer screening, and some of the aspects are indeed specific to cancer. However, most of the above reasoning applies also to screening for non-cancerous conditions. This is because the underlying premises are similar when screening asymptomatic people, no matter if the condition screened for is cancer or some other none-communicable disease or risk factor. To find an asymptomatic condition has in general no benefit in itself; the benefit comes from avoiding the symptomatic end-point. For people who would never experience the symptomatic end-point, a diagnosis of an asymptomatic condition is not beneficial in terms of disease outcomes, but can cause harm through unnecessary treatment and disease labelling (Welch 2011).

For example; an abdominal aortic aneurysm is defined as an aortic diameter equal to or above 30 mm (Wanhainen 2011). The risk of rupture is very small if the aneurysm is 30 mm, but increases with increased diameter of the aneurysm (Brown 2003). However, some small aneurysms do rupture while some large aneurysms never rupture. Additionally, aneurysms grow at different rate and up to half of all small aneurysms hardly grow at all (Vardulaki 1998, Thompson 2010). This means that abdominal aortic aneurysms have a non-predictable disease progression. Furthermore, we know from autopsy

studies that some people die with, and not from, un-ruptured abdominal aortic aneurysms (Bengtsson 1996), i.e. there is a disease reservoir for this condition. The same applies to asymptomatic cases of chronic obstructive pulmonary disease (Miller 2015), chronic kidney disease (Moynihan 2013), predementia (Le Couteur 2013), polycystic ovary syndrome (Copp 2017), hypertension (Martin 2014), diabetes type 2 (Cundy 2014, Yudkin 2014) and osteoporosis (Järvinen 2015). Arguably, all these asymptomatic conditions have disease reservoirs partly because they exist in a disease spectrum, not as either/or entities. They have non-predictable disease progressions because some people will suffer symptomatic consequences soon after diagnosis, while some will do so after an extensive time period, and further some would never suffer any consequences in their remaining lifespan even if the condition had remained undetected.

Furthermore, the same pattern as in the graph for thyroid cancer above, is observed for non-cancerous conditions. For example; the incidence of pulmonary embolism has risen substantially and simultaneously with an increased usage of computer tomography, while mortality has remained stable (Wiener 2013). This indicates non-predictable disease progression and a substantial disease-reservoir for this condition. In other words; the increased sensitivity of computer tomography compared to previous diagnostic techniques makes us find pulmonary emboli that do not need to be found (Wiener 2013).

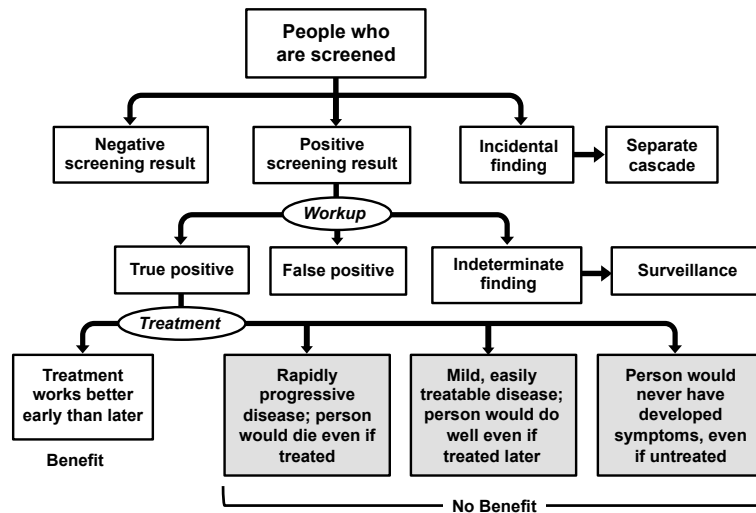
Indeed, an understanding of a diagnosis as something absolute is questionable. Instead of primarily focusing on whether a person fulfils the criteria for a certain diagnosis, it is probably more relevant to consider whether that person is likely to gain from receiving the diagnosis or not.

## Benefits and harms of screening - taxonomy

The general idea behind screening is to decrease mortality and morbidity through the detection and treatment of risk factors, precursors of disease or early stages of disease, in asymptomatic people. However, for screening to be effective it must not only detect disease at an earlier stage, but the early detection must also lead to improved prognosis, i.e. a lower incidence of late-stage, symptomatic disease and/or death (Welch 2011). Furthermore, if screening detects “disease” that would never have progressed to cause symptoms, the benefits of early detection might be outweighed by harms of over-detection of harmless conditions, and subsequent unnecessary medical interventions (Welch 2011, Harris 2014).

### **The screening cascade**

In order to think systematically about benefits and harms of screening, it is helpful to follow the steps in the screening cascade, shown on the next page in a figure from Harris and colleagues (on which the following section is based) (Harris 2014).



**Figure 3.** The screening cascade

From *Annals of Internal Medicine*, Harris RP, Sheridan SL, Lewis CL, Barclay C, Vu MB, Kistler CE, Golin CE, DeFrank JT, Brewer NT. The harms of screening: a proposed taxonomy and application to lung cancer screening. *JAMA Intern Med* 2014;174:281–5. American College of Physicians. All Rights Reserved. Reprinted with the permission of American College of Physicians, Inc.

### True negative results

People who are screened can have a negative or a positive result of the screening test. For those with a negative result, this may be either true or false (not shown in Figure above). For those with a true negative result, screening might offer reassurance, which can constitute a benefit from screening. However, it could also be questioned if such reassurance is beneficial in the long run, since it might create a need for recurrent reassurance and thus repeated medical investigations (Brodersen 2011). It may also lead to reluctance to seek medical attention for symptoms of a disease that arise after the screen (and thus poorer outcomes) (Goldenberg 2016).



### False negative results

For those with a false negative result (i.e. those who have a disease that the screening test do not find), screening might be harmful because having a negative screening test might reassure both the patient and health care personnel and the diagnosis thus might be delayed (Goldenberg 2016).

### False positive results

Having a false positive result means that the initial screening test is positive, but further work-up reveal that the initial finding was a false alarm (Harris 2014). This could be exemplified with a suspicious finding on mammography screening, which after ultrasound and biopsies is diagnosed as a benign tumour. Previously, false positive findings have been considered rather harmless, especially when the final diagnosis is derived within a short timeframe. However, in a large questionnaire study, women with false positive results on mammography screening reported greater negative psychosocial consequences, compared to women with normal findings, still three years after being declared free of cancer (Brodersen 2013).

### True positive results

Having a true positive result means having a positive finding on the screening test and actually having the condition. The diagnosis is most often confirmed only after further work-up, for example through histopathological investigation of a biopsy from a suspected lesion found at radiological investigation. People with true positive findings belong to one out of four categories, as displayed in the graph above (Harris 2014);

1. those for whom the earlier diagnosis improve prognosis,
2. those who would die of the disease at the same time with or without screening,
3. those for whom if the disease would have been detected first when it gave symptoms (i.e. without screening) it would still have been treatable with the same treatment and the patient would have survived the disease anyway,
4. those who would never have had any symptoms of the disease if it had not been detected through screening, i.e. they are overdiagnosed.

For people in category one, the early detection through screening is beneficial (Harris 2014). They can either avoid death of the condition screened for. This can be exemplified with a woman who has a non-metastatic screen-detected melanoma of the skin extirpated and that melanoma would otherwise not have been detected before metastases, which would have led to death. Or the early detection through screening can result in less aggressive treatment, which constitutes a benefit in itself (Harris 2014). This can be exemplified with a man who through preventive surgery of a screen-detected abdominal aortic aneurysm avoids a rupture of the aneurysm, which he would have survived also without screening. Since acute surgery for a ruptured aneurysm has a much higher complication rate than elective surgery, screening is likely beneficial for this man, although he would have survived also without screening.

For people in categories two and three, screening is not beneficial. Instead, screening results in living longer with a diagnosis without having ones life extended. This is considered a harm of screening since living with a history of a potentially life threatening disease can impact quality of life (Harris 2014). This can be exemplified by a man who gets an abdominal aortic aneurysm detected at screening at the age of 65. The aneurysm is small and grows slowly, and when the an-

eurysm is large enough for surgery to be considered at age 87, he is too old to be eligible for the operation. He dies from a ruptured aneurysm at the age of 89. For him, screening resulted in unnecessarily living with the knowledge of having an abdominal aortic aneurysm for 24 years, without any benefit.

People in category four are overdiagnosed. For them, screening results in harm through psychological consequences of being labelled with a diagnosis, and subsequent medical interventions, which per definition are unnecessary since the condition would never have caused symptoms (Harris 2014). Overdiagnosis was for long a controversial concept but is increasingly accepted as an important harm of screening (Welch 2011, Barratt 2015), although the magnitude of overdiagnosis caused by screening is still a topic of intense debate (Biesheuvel 2007). Overdiagnosis can be exemplified with a woman who, at the age of 62, is diagnosed with breast cancer at screening. She has surgery, and perhaps chemotherapy and radiation, for this breast cancer. She dies at the age of 85 of a heart attack, which she would have even if the breast cancer had never been detected, since this cancer would never have progressed to cause symptoms. Consequently, she was unnecessarily labelled with a diagnosis of breast cancer, and she unnecessarily received treatment for this cancer.

For screening programmes whose primary effect is finding and treating precursors of disease and through this lower overall disease incidence, overdiagnosis of the late-stage disease is generally not a substantial problem (Bretthauer 2013). In these cases, overdiagnosis and overtreatment of the precursor may be a bigger problem. For example; in cervical cancer screening, overdiagnosis of cervical cancer is probably limited, while overtreatment of dysplasia that would never have progressed to cause symptoms may be considerable (Moyer 2012).

## Incidental findings

For incidental findings, the screening cascade starts over again, i.e. the incidental finding could be a false positive or a true positive. People with true positive findings could either benefit from the early diagnosis, they could be “correctly” diagnosed but nevertheless not benefit from the early diagnosis, and they could be overdiagnosed (Harris 2014).

## Estimating benefits and harms of screening

For a screening programme to be worthwhile, it must not only have benefits, but the benefits must outweigh the harms. To be able to judge whether this is fulfilled or not, a prerequisite is that both harms and benefits are adequately explored, quantified and their consequences sufficiently investigated. To cover research methods for evaluation of all harms and benefits of screening is outside the scope of this thesis. I will now discuss methodological aspects related to estimating the effect of screening on mortality, overdiagnosis and overtreatment. I will briefly touch upon methods to estimate psychosocial consequences of screening.

### **Evidence from randomised versus non-randomised studies**

Before considering implementation of population-based screening in asymptomatic citizens, high-quality evidence from randomised trials showing a mortality benefit is a general requirement from official bodies like for example the World Health Organisation (Andermann 2008, UKNSC 2015). This is a legitimate requirement for many reasons. Firstly, beneficial effects of screening are small on the population level, and small effects require very high quality evidence to be revealed with confidence (Prasad 2016). Secondly, non-randomised studies of the effects of interventions on mortality have an inherently high risk of bias and can lead to seriously misleading results (Higgins 2011). Thirdly, screening always has harms, why there must be a high certainty of a benefit that may potentially outweigh these harms (Harris 2014). Fourthly, screening programmes has a high potential for opportunity costs (Harris 2014). Fifthly, when offering screening, health care systems invite asymptomatic people to an intervention

that they have not asked for, which leads to ethical considerations that differ from those in regular health care (Sackett 2002).

Consequently, evidence of a mortality benefit should, in general, come from well-conducted randomised trials before implementation of screening could come in question. Furthermore, the gold standard for estimating overdiagnosis and overtreatment is a high quality randomised trial with life-long follow-up, where the control group has never been screened (Carter 2015). However, some screening programmes have been implemented without evidence from randomised trials. As the introduction of screening in a population generally precludes additional randomised trials, evidence from non-randomised studies is usually needed in these cases. Furthermore, non-randomised studies are generally needed for continuous reassessment of existing screening programmes. Such reassessment is important because the premises of screening programmes can change after its introduction (Carter 2015). For example; the incidence of the disease can decrease or increase, the screening test can become more sensitive or the treatment of the disease can improve. Perhaps counter-intuitive, this last aspect may result in screening being less beneficial, since treatment might become equally effective for people with screen-detected disease as for those diagnosed later due to symptoms, making screening obsolete.

Consequently, we sometimes need to turn to non-randomised studies, such as cohort studies, for estimating the effect of screening on mortality as well as on overdiagnosis and overtreatment. A central problem in cohort studies is the risk of confounding, i.e. the difficulty of separating an effect of screening from an effect of other factors (Carter 2015). The choice of “control group” is crucial. The “control group” can consist either of historical cohorts (i.e. comparing trends before and after screening), of age groups not invited to screening (i.e. younger or older than those invited to screening) or of contemporaneous same-age cohorts not invited to screening (for example a

geographical region where screening has not been introduced) (Biesheuvel 2007, Carter 2015, Lee 2017). The last alternative is usually preferable since factors like time trends in mortality and incidence, and different trends in different age groups, do not affect such estimates. However, a situation that allows this require differential introduction of screening in various regions over a long time-frame, a condition that is only rarely met. Additionally, there is still a risk of confounding due to other factors such as differences in socioeconomic status between screened and non-screened groups that live in different geographical areas. Noteworthy, it is important not to use non-attenders as a control group, since non-attenders generally have a much higher overall risk of dying compared to attenders to screening, regardless of the effects of screening. In other words; attendance to screening is associated with a better health in general, a well-known phenomena called “the healthy screenee effect” (Raffle 2007). Therefore, such estimates will be biased in favour of screening.

In conclusion, evidence from non-randomised studies is needed to monitor the effect of existing screening programmes, which may have changed substantially since the trials that motivated them. But such studies have several methodological pitfalls, which should be carefully considered.

### **Estimating a mortality benefit**

Disease-specific versus total mortality

Disease-specific mortality in cancer screening trials is an outcome prone to bias from misclassification of the cause of death (Prasad 2016). Knowledge of the diagnosis increases the risk that the cause of death is falsely attributed to the disease in question (sticky-diagnosis bias) (Black 2002). Since, in general, more people in the screening

group are diagnosed with the condition compared to the control group, this bias will underestimate the effect of screening. Conversely, a death can be falsely attributed to another cause, usually because some time has elapsed since diagnosis or because the connection is not always clear-cut (slippery-linkage bias) (Black 2002). For example, a death due to renal failure arising after surgery for a screen-detected abdominal aortic aneurysm might not be attributed to screening, and likewise for a suicide due to psychological harms following the diagnosis of a screen-detected prostate cancer. Total mortality is free from these biases and is therefore the most reliable outcome when evaluating an effect of a screening programme (Prasad 2016). The downside of total mortality as an outcome is that large populations are needed in the trials to reliably detect a difference, especially when the effect is small in absolute numbers. As of now, no screening intervention has a documented effect on total mortality despite sometimes hundreds of thousands of randomised individuals (Prasad 2016).

#### Why survival rates are misleading

A common way of presenting the effect of screening is through survival-rates, i.e. the proportion of people diagnosed with the condition who is still alive X years after being diagnosed, in screened versus non-screened populations (Wegwarth 2012). Since the whole point with screening is to advance the diagnosis, survival rates will inevitably be improved for the screening group even if screening does not result in any survival benefit, i.e. even if all people affected by the condition die at the exact same time with or without screening (lead time bias). Additionally, in screening interventions with overdiagnosis survival rates in the screened group will be inflated, since overdiagnosed people per definition will not die from the condition. This leads to that screening appear more beneficial than it is (Wegwarth



2012). For example, in the case of neuroblastoma (referred to previously in this Introduction), screening increased survival rates from 17% to 72% (Sawada 1982), even though screening had no effect on mortality from neuroblastoma (Schilling 2002, Woods 2002). Consequently, survival rates are inherently biased and thus misleading in a screening context. They should not be used to compare screened versus non-screened groups (Wegwarth 2012).

### **Estimating overdiagnosis**

Overdiagnosis results in harm in two ways; through psychosocial consequences of being labelled with a diagnosis, and through subsequent medical investigations and treatments (Welch 2010, Welch 2011, Harris 2014, Barratt 2015, Carter 2015). These are per definition unnecessary since the condition would never have caused symptoms. It is not possible to know who is overdiagnosed on an individual level, but it is possible to estimate the rate of overdiagnosis caused by screening at the population level (Carter 2015).

It is important to realise that there is no single “correct” estimate of overdiagnosis for a given type of screening (Carter 2015). The level of overdiagnosis will change over time due to, for example, developments in underlying incidence rates and more sensitive screening technologies. There will also be differences between settings due to differences in diagnostic standards and practises. Any estimate of overdiagnosis is therefore a snapshot in time for a given context.

To conceptualise how overdiagnosis can be estimated, one could imagine a randomised trial with lifelong follow-up where people are randomised to screening or no screening. Typically, the incidence of the condition screened for will initially increase in the screening group compared to the control group. In the absence of overdiagnosis, this

initial increase will eventually be fully compensated for by a similar decrease in incidence in older age groups. Overdiagnosis due to screening is the absolute difference in the number of diagnoses detected during the lifetime of the two groups, provided the control group is not screened (Biesheuvel 2007, Zahl 2013).

#### Methods for estimating overdiagnosis

The best method to estimate overdiagnosis caused by screening is a topic of intense debate within the research community (Biesheuvel 2007, Zahl 2013, Lee 2017). Biesheuvel and colleagues summarise the different methods in a paper, which most of the following exposition is based on (Biesheuvel 2007). In summary, there are mainly two different ways of estimating overdiagnosis; the excess incidence approach (including the incidence-rate-method and the cumulative-incidence method) and modelling approaches (Biesheuvel 2007) (also called lead time approaches) (Etzioni 2015). Central to understanding the different methods is the concept of lead time. Lead time is the time by which screening advances the diagnosis, that is the time between detection at screening and the time when the condition would have presented itself clinically in the absence of screening.

There may not be one single “best method” to estimate overdiagnosis (Carter 2015, Etzioni 2015). This will certainly depend on the dataset available in a given setting (for example; is there a contemporary non-screened “control group” available or not). Indeed, as no method is perfect, the best approach may be to apply more than one method, each having its own strengths and weaknesses and using these to “triangulate” an estimate (Etzioni 2015).

## The excess incidence method

In the excess incidence method, incidence is compared between a screened and an unscreened group (average annual incidence in the incidence-rate-method and cumulative incidence in the cumulative-incidence method) (Biesheuvel 2007).

**The incidence-rate method** can be used in screening programmes where screening is performed repeatedly during an extended time period, like for example in breast cancer screening (Biesheuvel 2007). With the incidence-rate method, the annual incidence rate is compared between a screened and an unscreened group once a screening programme is well established. This means that, in addition to adjustments for lead time for incident cases, lead time is accounted for by excluding the early screening rounds. Excluding the initial screening rounds will however also exclude prevalent conditions that were never destined to become symptomatic, and this approach will thereby underestimate overdiagnosis (Biesheuvel 2007).

With **the cumulative incidence method**, the cumulative incidence of the condition screened for is compared between a screened and an unscreened group of people over the same time period (Biesheuvel 2007). Because lead time results in a drop in incidence after screening stops, the estimate should be performed first after screening stops plus the length of the maximal lead time (which is a topic of scientific controversy, see below). If there is no overdiagnosis due to screening, the cumulative incidence will be the same in the two groups after this time of follow-up. If there is overdiagnosis due to screening, there will be an excess of cases in the screened group compared with the unscreened group. Since this method includes the early screening rounds, it has been argued that it is more robust than the incidence rate method (Biesheuvel 2007).

The major criticism of studies using the cumulative incidence method

includes; that follow-up after screening stops is often too short (which will result in over-estimates of overdiagnosis), that continued screening in the screening group after screening stops is often unaccounted for (which will also overestimate overdiagnosis), and that estimates of counterfactual incidence (i.e. estimating what the incidence would have been without screening) are susceptible to considerable uncertainty in non-randomised studies, with potentially substantial impact on the estimates of overdiagnosis (Etzioni 2015, Lee 2017).

The modelling approach (also called the lead time approach)

While the excess incidence approach estimates overdiagnosis based on observed data, modelling approaches are based on modelling of disease transition (Biesheuvel 2007, Carter 2015). In this approach, estimates of overdiagnosis are based on a “competition” between death from other causes and lead time (i.e. overdiagnosis arises if the time of death comes before the time point at which the diagnosis would have presented itself clinically without screening). Central in this approach is therefore the estimate of lead time. The major criticism to modelling approaches is that they require assumptions, which are subjective and difficult, or even impossible, to verify. Furthermore, the models are also accused of a lack of transparency, i.e. it is difficult to check how the estimates have been derived (Biesheuvel 2007, Carter 2015). The pros of modelling approaches are that estimates of overdiagnosis can be derived even if empirical data from randomised trials or natural cohorts is lacking (Etzioni 2015, Lee 2017).

## The dispute about lead time

There is scientific controversy about appropriate methods to estimate lead time, i.e. the length of time that diagnosis is advanced by screening (Zahl 2013, Etzioni 2015, Lee 2017). Zahl and colleagues argue that a fundamental problem with most estimates of lead time is that they do not take overdiagnosis into account. Since overdiagnosed cases per definition have an infinite lead time, this will result in overestimates of lead time for the clinically relevant cases (Zahl 2013). An overestimated lead time for clinically relevant cases results in overcompensation, which in turn leads to underestimates of overdiagnosis (both for modelling approaches and for excess incidence approaches).

## Which denominator to use

When estimating the relative rate of overdiagnosis, the denominator could be either the conditions detected during the screening period, or the total number of conditions detected during the lifetime of the screening and control groups (Biesheuvel 2007). The first approach provides an estimate of the proportion of conditions that are overdiagnosed during screening, while the second approach estimates the life-time risk of being overdiagnosed due to screening. The second approach will lead to lower percentage estimates of overdiagnosis (absolute estimates of overdiagnosed cases are not dependent on a denominator). This is because the second approach is susceptible to dilution effect from new diagnoses that appear in both groups after screening. The first approach is arguably a more relevant measure, at least from the perspective of the individual deciding on whether to attend to screening or not (Biesheuvel 2007, Zahl 2013).

## **Estimating overtreatment**

When estimating overtreatment, much of the same principles mentioned above for overdiagnosis apply. But instead of incidence (i.e. diagnoses), the rate of treatment is analysed. Treatment could constitute pharmacological interventions, surgery or virtually any other medical treatment. Naturally, there are special considerations depending on the context.

## **Test accuracy of the screening test**

The test accuracy of a screening test is important when considering the benefits and harms of screening; a low sensitivity will result in a high rate of false negative results, while a low specificity will result in a high rate of false positive results – both affecting the benefit-to-harm balance of screening. However, even if the test accuracy of a screening test is good, it does not necessarily mean that this can be translated into a beneficial effect in patient relevant outcomes, i.e. that the screening test is good at finding the disease screened for does not necessarily mean that finding the disease through screening will improve prognosis (Hakama 2007, Hakama 2015). Further, measures of test accuracy do not normally take overdiagnosis into account, which means that in theory both sensitivity and specificity could be deceptively high even if most screen-detected conditions would constitute overdiagnosis (Hakama 2007, Hakama 2015). In conclusion, measures of screening and diagnostic test accuracy such as sensitivity and specificity could be seen as surrogate outcomes, which cannot necessarily be translated into patient relevant outcomes. An exposition of further aspects related to diagnostic test accuracy of the screening test is outside the scope of this thesis.

## Psychosocial consequences of screening

The concept of separating physical and psychosocial consequences of screening exposes a body-mind dualism inherent in our scientific paradigm (Kirkengen 2016). Can the consequences of going through massive surgery (with risks of major complications and even death) be separated into effects in our bodies and effects in our minds? Can the existential uncertainty following a diagnosis of a life threatening condition be defined as solely psychological in nature? Emerging evidence questions this dualistic perspective, and suggests that mind and body are not only two entities affecting each other, but so intertwined that they are basically one and the same (Kirkengen 2016). However, for reasons of simplicity, I here chose to adhere to the prevailing taxonomy of psyche and soma.

Shortage of evidence on psychosocial consequences of screening is common; a review found this for all screening modalities investigated: abdominal aortic aneurysm, prostate cancer, lung cancer, osteoporosis and carotid artery stenosis (DeFrank 2015). Most of the quantitative studies on psychosocial consequences of screening use generic (as opposed to diagnosis-specific) questionnaires (DeFrank 2015). Examples of generic questionnaires are SF-36, ScreenQL, EQ-5F and HAD. Generic questionnaires have a low validity in a screening context because they do not capture central aspects specific to the screened condition (McCaffery 2004, Brodersen 2007), for example anxiety about rupture during sexual activity for people with screen-detected abdominal aortic aneurysms. Also, aspects not important for this specific group can contaminate the results. The use of generic instruments is therefore questioned in a screening context (McCaffery 2004, Brodersen 2007). Additionally, the psychometric properties of the used questionnaires in studies of psychosocial consequences of screening are often not reported, which arguably signals a low quality of these studies (Brodersen 2007). However, methodological aspects

of estimating psychosocial consequences of screening are outside the scope of this thesis. Noteworthy, some aspects of being labelled with a diagnosis are not easily captured through quantitative approaches, why it is problematic that qualitative research is generally considered less important than evidence from quantitative approaches in evidence synthesis and policy making (Greenhalgh 2015).



# Beyond benefits and harms

## **Costs, feasibility and equity**

For a screening programme to be worthwhile, it must not only have an acceptable benefit-to-harm balance. It must also be cost-effective and feasible (Andermann 2008). Additionally, aspects of equity should be considered, as well as opportunity costs (Andermann 2008). An in-depth analysis of these factors is outside the scope of this thesis, why I will only briefly consider some of the related aspects.

In most screening programmes, attendance displays a social gradient (Weiss 1996, Raffle 2007), which means that a higher proportion of people with high economic status attend to screening compared to people with low socioeconomic status. In this way, screening programmes often contribute to inequity in health care. Ironically, for most conditions screened for, people with low socioeconomic status have a higher risk of dying from the condition compared to people with high socioeconomic status (Weiss 1996, Raffle 2007).

Most cost-effectiveness analyses of screening does not adequately include costs from important harms, which make them misleading. Additionally, screening programmes for asymptomatic conditions might influence the general perception of disease and health in a society; our ability to trust our own bodies might be affected by being diagnosed with disease while being free of symptoms (Reventlow 2006, Sångren 2009, Harris 2014). In extension, this might affect search patterns for health care services and would in that case have economical effects outside of the screening cascade. Furthermore, screening results in opportunity costs (Harris 2014). This means that resources are spent on screening that could have been spent on other medical interven-

tions, or in other societal sectors.

### **The ethics of screening**

Screening will always cause harm to some people, even when resulting in substantial benefit for other people (Shickle 1994, Harris 2014). There is no scientifically correct answer to whether this is ethically acceptable. From a utilitarian perspective, screening is justifiable if it results in net benefit (Shickle 1994, Kelly 2015). However, science cannot provide an answer on what constitutes a net benefit; this is a value judgement (Harris 2015, Kelly 2015, Carter 2017). Moreover, strict utilitarianism is hardly acceptable in Western medicine, we must also consider overarching deontological principles, such as the general requirement for the medical profession of not causing harm (*primum non nocere*), especially when inviting asymptomatic citizens to an intervention they have not asked for (Shickle 1994).

As more attention is given to the harms of screening, it is increasingly argued that the solution to the ethical dilemmas presented by screening is to ensure “free and informed decisions of those who are invited to screening” (Brownsword 2010), which correlates to an increased focus on informed choice and shared decision making in health care at large (Hoffmann 2014). But many screening programmes do not provide information about harms in invitations (Jørgensen 2009, Gøtzsche 2011, Kolthoff 2016), which overrides the autonomy of the individual. Further, a scarcity of reliable information about the harms of screening in general means informed decisions without adequate access to the facts.

### **When is a screening programme worthwhile?**

In conclusion, a prerequisite for being able to judge if a screening programme is worthwhile is that both benefits and harms are adequately investigated and that aspects of costs, feasibility, equity and ethics have been carefully considered. However, there is no scientifically “correct” way of judging whether a screening programme is worthwhile or not (Harris 2015, Carter 2017). There is no common unit of measurement for benefits and harms of screening why the balance will always be a subjective judgement, which no one has exclusive privilege to assess. Therefore, when authorities decide if a screening programme should be implemented or not, or if an existing screening programme should be deimplemented or continued, the process should be transparent and the inherent value judgements should be made explicit (Barratt 2017, Carter 2017).

I will discuss themes related to opportunity costs, equity and the ethics of screening in more depth in the General Discussion of this thesis.

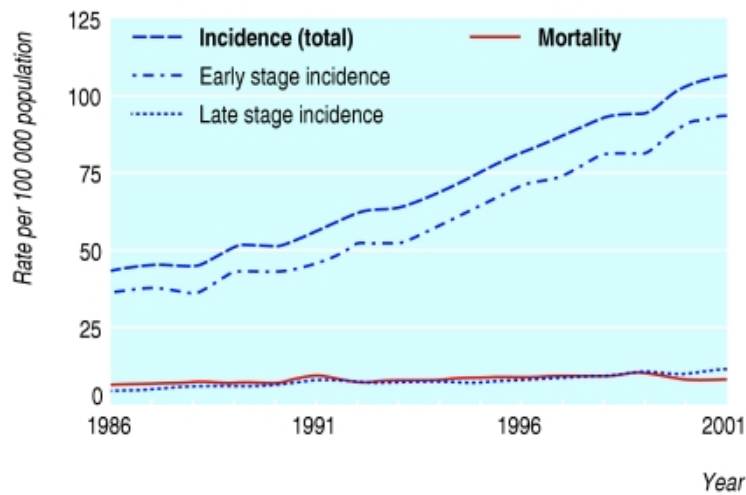
## Screening for Malignant Melanoma

Malignant melanoma is a tumour of the skin, which can cause death through metastases to other organs. The most important avoidable risk factor is irregular over-exposure to ultraviolet radiation from sunlight and artificial sources (Gandini 2005). Other risk factors include blonde or red hair, green or blue eyes, freckles, an inability to tan, a family history of malignant melanoma, and a large number of naevi and dysplastic naevi (Marks 2000).

Screening for malignant melanoma has the potential to reduce mortality from the disease through earlier detection, as prognosis is closely associated with the thickness of the lesion at the time of diagnosis, with thinner lesions having a much lower risk of metastases (Breslow 1970). Screening for malignant melanoma can be performed through visual self-examination of the skin or visual inspection by a general practitioner, dermatologist, or other health professional, which can be followed by dermatoscopy of identified lesions. Other methods to assist in diagnosing malignant melanomas are evolving and might also be used for screening, for example, teledermatology, mobile phone applications, and spectroscopy-based techniques (Dinnes 2015).

The incidence of malignant melanoma in Western populations has risen many-fold over recent decades (Garbe 2009). This might be due in part to an increase in exposure to risk factors, mainly ultraviolet radiation from the sun and artificial sources (Waldmann 2012). However, it is also likely that some of the rise in incidence is caused by overdiagnosis of indolent malignant melanomas through increased disease awareness and screening. As displayed in the graph below from Welch and colleagues, the large increase in incidence has not been followed by an increase in mortality (Welch 2005). As described previously, such patterns are unlikely to be explained by improve-

ments in treatment (for example an increased rate of extirpations in the case of melanoma), since such effects would have to simultaneously and exactly equal out the increase in incidence throughout an extended time period. Instead, the observed pattern is suggestive of a large rate of detection of melanomas that do not need to be found, i.e. overdiagnosis (Welch 2011).



**Figure 4.** Incidence and mortality from malignant melanoma in a United States population aged 65 and older, 1986-2001. Early stage refers to in situ and local disease; late stage refers to regional and distant disease.

Reproduced with permission from (Welch HG, Woloshin S, Schwartz LM. Skin biopsy rates and incidence of melanoma – population based ecological study. *BMJ* 2005;331:481) Copyright © 2005, British Medical Journal Publishing Group.

Screening for malignant melanoma is not recommended in the United States (Wernli 2016), Canada (CTFPHC 2013), Australia, or New Zealand (ACNMGRWP 2008). Germany has had a national screening programme for malignant melanoma since 2008 (Katalinic 2015) and

opportunistic screening is increasingly used in many Western countries (Lakhani 2014). In Australia, the annual skin screening rate range from 10% to 50% of the adult population depending on how skin screening is defined (Girgis 1991, Balanda 1994, Heywood 1994, Borland 1995, Janda 2004), and the corresponding rate in the US is 14% to 20% (Federman 1997, Ford 2004, Sarayia 2004, Federman 2006). Several professional societies, who may have inherent vested interests, recommend skin screening (Jørgensen 2017). In Europe, a campaign involving dermatologists in over 30 countries (EUROMELANOMA) recommends "*visiting your dermatologist regularly for a skin check-up*" and conducting self-examination every month (EADO 2016). In the US, the American Cancer Society recommends a skin self-exam every month (American Cancer Society 2017) and the American Academy of Dermatology runs a skin screening programme wherein over 2.5 million skin screens have been conducted since 1985 (AAD 2017).

Consequently, screening for malignant melanoma is currently practised in many countries, apparently without support from randomised trials. This is problematic since screening for malignant melanoma likely causes overdiagnosis of harmless malignant melanomas and subsequent overtreatment (Welch 2005, Norgaard 2011). In addition, screening for malignant melanoma has a high potential for opportunity costs. In conclusion, it is essential to evaluate the evidence base for benefits and harms of screening for malignant melanoma.

## Screening for Abdominal Aortic Aneurysm

Abdominal aortic aneurysm (AAA) is a widening of the abdominal aorta and is defined as an aortic diameter equal to or above 30 mm (Wanhainen 2011). AAAs are usually asymptomatic until they rupture, which is fatal in more than 80% of cases (Basnyat 1999). Risk factors for developing AAA are smoking, male sex, advanced age, and family history of AAA (Guirguis-Blake 2014). The risk of rupture is correlated to the size of the aneurysm (Brown 2003). According to guidelines, men with aneurysms of 30 to 54 mm are offered regular ultrasound surveillance for the rest of their lives. If the aneurysm is equal to or above 55 mm or grows more than 10 mm annually, elective surgery is considered (Cosford 2007). Screening aims to detect the aneurysm before it ruptures, enabling elective surgery and hence reducing morbidity and mortality from rupture.

Elective surgery for AAA is associated with a risk of mortality and serious complications such as impotence, myocardial infarction, stroke, amputation, respiratory failure, renal failure, ischaemic colitis, spinal cord ischaemia and prosthetic graft infections (Calonge 2005, Pettersson 2009, Linné 2014). A systematic review concluded that 30-day mortality for open repair was 4.7% (Stather 2013). Endovascular techniques reduce 30-day mortality, and the same systematic review reported a mortality rate of 1.3% for endovascular repair. However, endovascular techniques have more long-term complications and mortality is similar to that for open surgery after two years (Stather 2013). Mortality rates may have been lowered by improvements in operative techniques since the studies included in the above-mentioned systematic review. Screen-detected AAAs may have a lower perioperative risk. However, studies have not been able to confirm this (Guirguis Blake 2014, Linné 2014) and one of few studies on the topic found that 44% of all screen-detected AAAs operated on was

defined as “complex” from a surgical perspective, which is comparable to the rate amongst incidentally detected AAAs (Ohlsson 2016).

The psychosocial consequences of living with an AAA under surveillance are poorly investigated (Guirguis Blake 2014, DeFrank 2015). Qualitative studies indicate important problems and living with the knowledge of having an AAA has been described as living with a “ticking bomb inside your stomach” (Hansson 2012). However, results from qualitative studies do not address the magnitude of these problems. The available quantitative studies all use generic (as opposed to diagnosis-specific) questionnaires (DeFrank 2015), which have a low validity in a screening context because they do not capture central aspects (for example fear of having sex because of fear of rupture) and since aspects that are not related to screening contaminate the results (McCaffery 2004, Brodersen 2007, DeFrank 2015). Additionally, none of the quantitative studies on the psychosocial consequences of living with the knowledge of having an AAA report on the statistical psychometric properties of the used questionnaires, which arguable signals a low quality.

Population level screening for AAA was introduced in Sweden (Wanhainen 2011), the United States (Guirguis-Blake 2014), and the United Kingdom (Davis 2013) during the 2000s. In Sweden and the UK, all men aged 65 are offered screening with a one-off ultrasound examination (Wanhainen 2011, Davis 2013). In the US, screening is recommended for men aged 65 to 75 who have ever smoked (Guirguis-Blake 2014).

The decision to implement screening was based on four randomised trials performed in the 1980s and 1990s. These trials were included in a systematic review from the US Preventive Services Task Force (USPSTF) from 2014, which concluded that AAA screening results in about 50% relative reduction in AAA-related mortality (Guirguis-Blake 2014), which translates into a 0.5 percentage point disease spe-



cific absolute mortality reduction. However, since the USPSTF review the long-term follow-up of one of the trials have been published and showed no beneficial effect (McCaul 2016). Consequently, the randomised trials collectively display great heterogeneity in their effect estimates, from a large beneficial effect in relative numbers in two trials (Lindholt 2006, Thompson 2012), to no effect in two trials (Ashton 2007, McCaul 2016).

Furthermore, these estimates were based on populations with a much higher disease incidence than today; in the UK, the incidence of AAA amongst 65-year old men has fallen by over 70% during the last decades (Choke 2012, Darwood 2012). Similar trends have been observed in Sweden (Svensjö 2011). This is likely caused by reduced smoking; in 1974, 45% of the population in the UK were smoking compared to 20% in 2012 (Office for national statistics 2012). A reduced incidence of the condition screened for results in reduced absolute benefit and possibly a less favourable benefit-to-harm balance.

To lower the threshold for the diagnosis of an AAA from 30 to 25 mm is advocated for in the scientific community (Svensjö 2011, Darwood 2012, Thompson 2012) and this has been partly adopted in the Swedish screening programme, which could further alter the balance of benefits and harms of AAA screening.

There is limited evidence on AAA screening for women (Cosford 2007, Guirguis-Blake 2014). Women have a much lower AAA incidence than men (Svensjö 2013) and women generally die of AAA at a higher age than men (Vänni 2016). A lower incidence of the condition screened for results in less benefit in absolute numbers. Additionally, a shorter expected remaining lifespan results in less life years gained from screening. Furthermore, surgery for AAA has a higher mortality rate in older age groups, as well as in women compared to men for all age groups (Ulug 2017), why the harms of screening might

be higher in women. Consequently, the benefit-to-harm balance from screening women for AAA is likely to be worse than for men.

Before the start of the studies included in this thesis, there were already systematic reviews on AAA screening from Cochrane (Cosford 2007) and from the USPSTF (Guirguis-Blake 2014). However, neither of these included estimates of overdiagnosis caused by AAA screening, nor analyses of the effect of a lowered threshold for an AAA-diagnosis on the rate of overdiagnosis. Before the studies included in this thesis there were, to our knowledge, no published estimates of the rate of overdiagnosis caused by AAA screening.

Furthermore, due to the radical drop in AAA-incidence the randomised trials of AAA screening are out-dated, which means that the benefits and harms of AAA screening in contemporary populations are unknown.

# Aims

The general aim of this thesis was to explore how benefits and harms of screening for non-communicable, potentially life-threatening disease can be evaluated, with a focus on challenges when up-to-date data from randomised trials is lacking.

## Papers I and II

- To assess the benefits and harms of screening for malignant melanoma in the general population in a systematic Cochrane review of randomised trials.

## Paper III

- To assess the evidence base for harms of screening for abdominal aortic aneurysm.
- To estimate the rate of overdiagnosis and overtreatment caused by screening for abdominal aortic aneurysm based on available data from previous studies.
- To estimate the effects of a lowered threshold of the diagnosis of an abdominal aortic aneurysm on the rate of overdiagnosis caused by screening for abdominal aortic aneurysm.

## Paper IV

- To estimate the effect of organised screening for abdominal aortic aneurysm in Sweden, in a population with low disease-incidence compared to the randomised trials that led to the introduction of screening, on overdiagnosis, overtreatment and disease-specific mortality.



# Methods

	<b>Design</b>	<b>Method</b>	<b>Data</b>	<b>Analysis</b>
<b>Paper I</b>	Protocol for Systematic Review	Standard Cochrane Methods	Protocol	Protocol
<b>Paper II</b>	Systematic Review	Standard Cochrane Methods	Data from RCTs (post-hoc from non-randomised studies)	Qualitative synthesis (no meta-analyses performed due to lack of data)
<b>Paper III</b>	Reanalyses and narrative review	Assessment of evidence base for harms and reanalyses of data from published studies	Published data from RCTs and non-randomised studies	Excess incidence method and qualitative synthesis
<b>Paper IVa</b>	Registry study - cohort	Analyses of differences between a cohort of screened versus a cohort of non-screened men	Individual data from registries: cause of death, inpatient, outpatient and Swedvasc, socioeconomic data from Statistics Sweden	Unadjusted - Nelson-Aalen plots (incidence rate ratios, risk differences) Adjusted - propensity score weighting, generalized estimating equations (odds ratios, risk differences)
<b>Paper IVb</b>	Registry study - before and after	Trends before and after screening and in screened versus non-screened age groups	Ecological data from registries: cause of death, inpatient, outpatient and Swedvasc	Descriptive statistics

**Table 1.** Summary of methods of the included studies

## Papers I and II

We performed a Cochrane systematic review of benefits and harms of screening for malignant melanoma. Prior to performing the systematic review, we published a peer-reviewed protocol where all outcomes, methods and statistical analyses were pre-specified. We used standard methodological procedures expected by Cochrane (Higgins 2011).

Our inclusion criteria were: randomised controlled trials, including cluster-randomised trials, that compared screening for malignant melanoma with no screening, regardless of screening modality or setting, in any type of population and in any age group where people are not suspected of having malignant melanoma.

## Paper III

We conducted a narrative review of the available evidence base for harms of AAA screening. We included data from the randomised trials of AAA screening found in the systematic review on AAA screening from the USPSTF. We also included data from studies not included in the USPSTF review: cohort studies of AAA screening, observational studies of trends in AAA-incidence and qualitative studies on psychosocial consequences of AAA screening. We established definitions of overdiagnosis and overtreatment in AAA screening. We estimated rates of overdiagnosis and overtreatment caused by AAA screening, as well as the effects of a lowered threshold for the AAA-diagnosis on the rate of overdiagnosis, based on published data from 1) randomised trials and 2) a cohort study, using the excess-incidence approach.

## Paper IV

The AAA screening programme in Sweden was introduced step-wise county by county between 2006 and 2015. Together with the well-developed population registries in Sweden, this presented a unique possibility for the evaluation of a contemporary, public AAA screening programme.

For study IVa, we conducted a cohort study using individual, anonymised registry data on AAA-incidence, the rate of AAA-surgery and AAA-related mortality between 2006 to 2015, for a cohort of 25 265 men invited to screening within the Swedish AAA screening programme and a contemporaneous cohort of 106 087 age-matched Swedish men not invited to screening because they lived in geographical regions where screening was not yet implemented. We analysed differences in the outcomes between the screening cohort and the non-screening cohort at 6 years of follow-up.

For study IVb, we analysed trends for the same outcomes before and after the introduction of screening, and in screened versus non-screened age groups, in all Swedish men 40 to 99 years of age between 1987 and 2015.

The registries used in study IVa and IVb were the Swedish Cause of Death Registry, the Swedish inpatient and outpatient registries and the Swedish national registry for vascular surgery (Swedvasc). Individual data on socioeconomic status and emigration for the men included in our cohorts, as well as population statistics for all men above 40 years of age, were retrieved from Statistics Sweden.





# Results

## Papers I and II

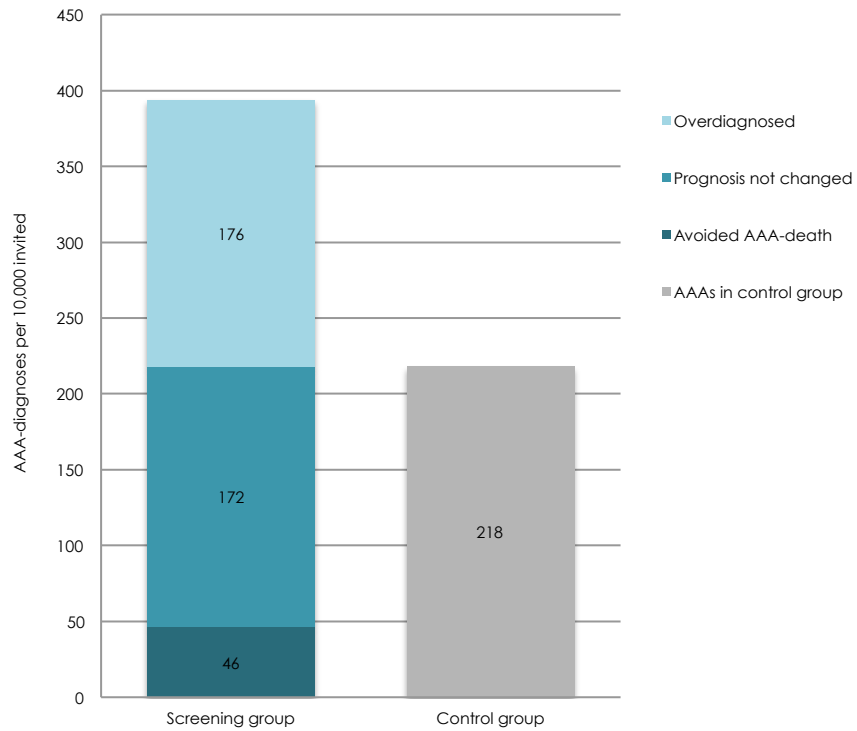
Two studies met our inclusion criteria, but these were not designed to estimate our outcomes; the first study was a randomised trial of an intervention developed to increase performance of skin self-examination and the second study was a pilot-study for a planned cluster-randomised trial of population-based screening for malignant melanoma. No results that could be included in our analyses were reported from these studies. At the time of the review, there were no published or unpublished data available on our pre-specified outcomes.

Post-hoc, we choose to evaluate some of the most influential non-randomised studies suggesting a beneficial effect of screening for malignant melanoma. We found a serious risk of bias in these studies.

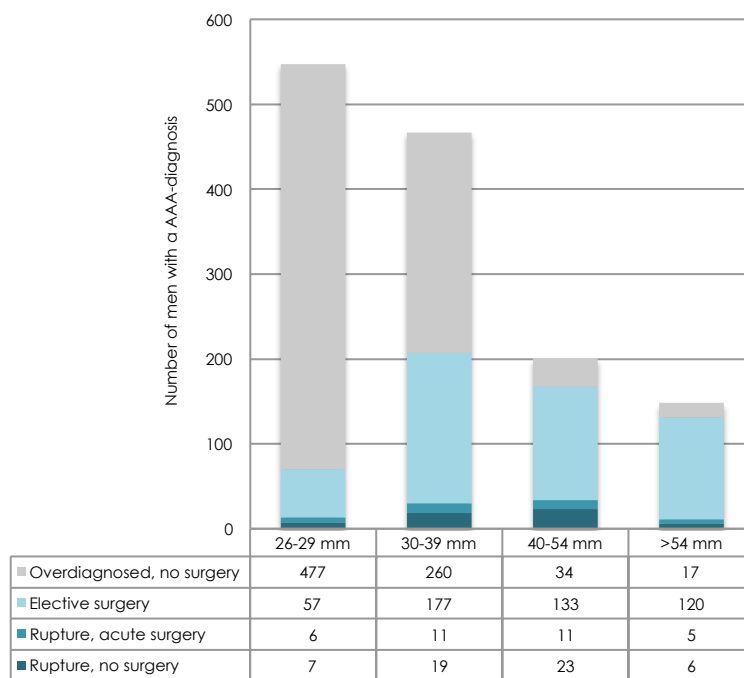
## Paper III

We found no previous estimates of overdiagnosis caused by AAA screening. Based on data from the randomised trials performed in the 1980s and 1990s, we estimated that 176 of 10 000 men invited to AAA screening were overdiagnosed with AAA 13 years after screening (95% CI 150 to 202) (Figure 5). These men were unnecessarily turned into patients and may have experienced appreciable anxiety throughout their remaining lives. Moreover, 37 of these men unnecessarily had preventive surgery (95% CI 15 to 60) and statistically 2 of

them died as a consequence. According to our estimates, 45% of all AAAs detected at screening were overdiagnosed 13 years after screening. Based on data from a cohort study with 10-20 year follow-up, we estimated that lowering the AAA definition from 30 to 25 mm would double AAA prevalence and substantially increase the rate of overdiagnosis (Figure 6).



**Figure 5.** Overdiagnosis estimated from a randomised trial with 13-year follow-up. Overdiagnosis was estimated to 45% of all screen-detected abdominal aortic aneurysms.



**Figure 6.** Overdiagnosis according to diameter at initial scan estimated from a cohort study with 10-20 year follow-up. Overdiagnosis rates: 26-29 mm: 87%, 30-39 mm: 56%, 40-54 mm: 17%, above 54 mm: 11%, all above 30 mm: 38%, all above 26 mm: 58%.

## Paper IV

In our analysis of population trends (IVb), we found that AAA-mortality has dropped about 70% in Swedish men aged 65 to 74 over the past two decades. Reductions were similar in areas offered and not offered screening, and in screened and non-screened age groups. In our analysis of screened and non-screened cohorts (IVa), we found

a non-significant reduction in disease-specific mortality with screening 6 years after screening (odds ratio 0.76, 95% CI 0.38 to 1.51). In absolute numbers, this means 2 men (95% CI -3 to 7) may have avoided AAA-death for every 10 000 men invited (not statistically significant). At the same time, 49 men were likely overdiagnosed (95% CI 25 to 73), of whom 19 men were likely overtreated, i.e. had unnecessary surgery with a risk of mortality and morbidity (95% CI 1 to 37). The remaining 30 men were offered regular follow-up with potential psychosocial consequences.

# Discussion of the studies

## Papers I and II

Screening for malignant melanoma has been adopted in many Western countries although there is no evidence from randomised trials of a beneficial effect. The most influential non-randomised study suggested a beneficial effect of screening for malignant melanoma but had a serious risk of bias favouring screening. Additionally, other non-randomised studies suggest harmful effects due to substantial overdiagnosis as well as considerable opportunity costs (Welch 2005, Norgaard 2011).

### **Limitations**

Our choice to include only randomised trials in this systematic review entailed some limitations. Our rationale was that including data from non-randomised studies in a Cochrane review on population screening might legitimise the use of such interventions based on low-quality evidence for a mortality benefit. However, if the merits of studies that have been pivotal for far-reaching public health decisions are not assessed, it reduces the relevance of the corresponding Cochrane reviews. As a compromise we choose post-hoc to evaluate the most relevant non-randomised studies in the Discussion section of the review. However, to evaluate some non-randomised studies without performing a systematic search might have introduced selection bias.

Furthermore, to rely on evidence from randomised trials only when evaluating benefits and harms of screening is undoubtedly problematic, since many of the effects of screening are not easily captured by randomised trials (see the General Discussion). The reason that this strategy might be motivated here is that arguably screening should not even be considered if there is no evidence from high-quality randomised trials of a beneficial effect on mortality; if there is no reliable evidence of a benefit, then screening asymptomatic citizens is not warranted, no matter what the harms are (Andermann 2008).

## Paper III

Screening for abdominal aortic aneurysm was introduced in several countries without adequate exploration and quantification of harms. Our study suggests that AAA screening causes considerable harm through overdiagnosis and overtreatment. Furthermore, we found that the rate of overdiagnosis would substantially increase if the cut-off for the diagnosis of an AAA was lowered from 30 to 25 mm. Noteworthy, this change of definition of an AAA is advocated for in the scientific community, and has been partly adopted within the Swedish screening programme, although the potential harms have not been adequately investigated. Moreover, the recent drop in AAA-incidence reduces the absolute benefit of screening and might worsen the benefit-to-harm balance, which means that data on both benefits and harms from the randomised trials of AAA screening is out-dated.

### **Limitations**

One limitation to this study is the non-systematic approach. This was based on pragmatism; there were already systematic reviews of this topic from Cochrane (Cosford 2007) and the USPSTF (Guirguis-

Blake 2014). It was not likely that we would have found randomised trials that they had not. The aim of this study was to explore harms that had not been adequately addressed in these reviews, which might motivate our non-systematic approach. However, there is a risk of selection bias. Furthermore, to lump results from the randomised trials of AAA screening might be questionable, since they display considerable heterogeneity. We did not explore this heterogeneity, which is a methodological weakness of our study. Regarding our estimates of overdiagnosis and overtreatment, there were multiple biases due to inherent limitations of the available data. These are described in the publication.

## Paper IV

The substantial changes in AAA-incidence and AAA-mortality for reasons other than screening motivate contemporary estimates of the effects of the intervention. The gradual implementation of AAA screening in Sweden and the availability of reliable population data presented a unique possibility.

Screening only had a minor effect on the decline in AAA-mortality in the Swedish population. The observed large reduction at population level was likely caused by reduced smoking. The effect on AAA-mortality in Sweden was 7% of that in the largest randomised trial at similar follow-up (2 vs 27 avoided AAA-deaths per 10 000 invited men). The observed rate of overdiagnosis was 28% of that estimated from the same trial (49 vs 176 per 10 000 invited men) and the rate of overtreatment was 51% (19 vs 37 per 10 000 invited men). This means that the benefit-to-harm balance is substantially worse today compared to that seen in the randomised trials that led to the introduction of AAA screening.

An aspect that we did not investigate in our study is that AAA screening seems to result in unforeseen diagnostic cascades, with potentially substantial effects on the benefit-to-harm balance of the intervention. Within the Swedish AAA screening programme, additional screening for iliac aneurysms is regularly performed (personal communication – mail correspondence available on request) and men with screen-detected AAAs are routinely screened for popliteal aneurysms (Wrede 2017). A study from Sweden showed that 24% of all men undergoing surgery for popliteal aneurysms were diagnosed due to AAA screening, a number that is likely to rise in the future (Wrede 2017). Thoracic or thoracoabdominal aneurysms are sometimes detected at angiography of the aorta performed due to a screen-detected AAA (Ohlsson 2016). Some researchers even advocate that all patients with an AAA should also be screened for thoracic aneurysms (Ziganshin 2016). The benefits and harms of such additional screening are not adequately investigated. The potential for harm is appreciable since elective surgery of asymptomatic aneurysms has considerable risks – in the case of thoracic aneurysms mortality is much higher than for abdominal aneurysms. Furthermore, one study found that 59% of patients undergoing elective AAA-surgery with endovascular techniques had an aortic diameter below 55 mm (Schanzer 2011) despite evidence that surgery on small AAAs is not beneficial (Guirguis Blake 2014). It has been argued that AAA screening, by detecting many small AAAs, is a major driver behind this development (Harris 2012).

These examples demonstrate that the consequences of screening programmes in a “real life” setting are difficult to foresee. There are often downstream effects that have not been investigated in the randomised trials. Arguably, the above-mentioned sequels of AAA screening should be carefully evaluated and taken into account in both cost-effectiveness analyses and estimates of the benefit-to-harm balance of AAA screening in the future.



In conclusion, the benefit-to-harm balance of AAA screening in contemporary populations is questionable and our study suggests that the continued justification of the screening programme should be revisited. However, neither new evidence of harms of AAA screening, conflicting results from long-term follow-up of randomised trials, nor a dramatic drop in disease-incidence substantially limiting the beneficial effects of AAA screening have led to consideration about the continuation of the screening programmes (Keller 2016a, Keller 2016b, Larsson 2016). Instead, implementation of screening for abdominal aortic aneurysm is discussed in additional countries (Lindholt 2013). Further, expanding screening by including women, despite lack of evidence (Ulug 2016), and by lowering the cut-off diameter for the diagnosis is advocated for in the scientific community (Svensjö 2011, Darwood 2012, Thompson 2012) (related papers 2 and 3 in Appendix).

### **Limitations**

This is a retrospective registry-based cohort study. Although we have a contemporaneous, same-age control group of non-screened men, it is not a randomised trial and there are differences in socioeconomic factors between compared groups, as well as differences between regions in AAA-mortality and incidence that pre-date organised AAA screening. Although we did adjust for socioeconomic factors on an individual level, we cannot exclude residual confounding. Further, there is a risk of misclassification in the registries, which is discussed in more detail in the publication. It is not possible to know if, or in what direction, this may introduce bias in our estimates.

Even though our hypotheses and analysis plan arose before we obtained data, we did not publish a pre-registered protocol, making selective outcome reporting impossible to control for by independent

researchers. Furthermore, due to restrictions from the registries and ethical committee, we cannot share individual data, making independent reanalysis impossible, except for analyses of aggregated data.

Since meaningful follow-up was limited to 6 years in our study the full effect of screening might not have been captured, which may lead to an underestimate of the benefit and overestimates of overdiagnosis and overtreatment. In the largest randomised trial of AAA screening, the majority of the effect on AAA-mortality was obtained at 7 years; 0.27 percentage point reduction in AAA-mortality at 7 years (Kim 2007), 0.42 At 10 years (Thompson 2009), and 0.46 at 13 years (Thompson 2012). If applying trends from this trial to our data, the absolute effect of screening on AAA-mortality would increase from 0.02 percentage points at 6 years to 0.03 at 13 years of follow-up. Even if some degree of overestimation of overdiagnosis and overtreatment cannot be excluded, we do not believe that further follow-up would change the overall conclusion of our study.

Despite thorough attempts, we were unable to obtain basic information such as precise starting dates and age groups invited from those responsible for the Swedish screening programme. Judging from incidence peaks observed in our dataset, we found indications that men may have been screened systematically prior to their 65<sup>th</sup> birthday, leading to underestimates of overdiagnosis and overtreatment in our analyses. Indeed, it is regrettable that outside evaluation of the screening programmes is associated with such hassles.

# General discussion

Diseases like malignant melanomas and abdominal aortic aneurysms can have devastating consequences. Numerous people have seen their loved ones depart this life too early in the trail of aggressive, fatal disease. We all share the vision that medicine will one day find ways to avoid such suffering and loss. Indeed, screening has a potential to cause a lot of good. However, the envisioned benefits are not always materialized in praxis, and mass screening seems to have an inherent potential to cause unintended harms.

The main overarching finding of this thesis is that benefits of screening receive much more attention and appreciation than harms in the scientific literature. This is not a novel finding, nor specific to the screening modalities investigated in this thesis (Jørgensen 2007, Jørgensen 2009, Woloshin 2012, Heleno 2013, Harris 2014, DeFrank 2015, Kolthoff 2016).

Arguably, screening for malignant melanoma has been adopted in the Western world based on an unfounded enthusiasm for screening, likely underpinned by vested interests (Jørgensen 2017). In the case of screening for abdominal aortic aneurysm, it seems as if a disease-specific mortality benefit seen in two of the four randomised trials, although out-dated and small in absolute numbers, trumps most counterarguments when it comes to policy making. In this context, it is important to reflect on where the burden of proof lies: no matter how terrible the disease in question is, we should be reasonably certain that a screening programme does more good than harm before implementing it. This thesis is devoted to that task.

I argue that the cases of screening for malignant melanoma and screening for abdominal aortic aneurysm point to some of the fundamental problems inherent in medical thinking, science and policy-making today. The picture on the front-page of this thesis is meant to illustrate this viewpoint. It is a reference to “the streetlight effect”, i.e. when we search for our keys where there is light, although we dropped them somewhere else. This is a metaphor for our propensity to focus on aspects that appear well demarcated and intuitively appealing to seek control over, while ignoring less approachable but sometimes more meaningful answers that are only to be found in-between the biomedical light posts.

In this General Discussion I will apply a wider perspective and contemplate on the broader and deeper meaning of the topic of this thesis; screening for potentially fatal disease. Never abandoning my respect for people who might have good reasons for seeing things differently than me, this General Discussion will focus on the problematic aspects of screening.

Some of the reasoning below has been published by me and my co-authors in papers related to this thesis (see List of related papers). These are available in the Appendix.

## The fundamental appeal of screening

Why is screening so popular and un-problematized in Western societies? The enthusiasm does not seem to be proportional to the factual evidence since the documented effect of even the most beneficial screening programmes is quite small in absolute numbers (Prasad 2016). Indeed, far more effective interventions receive much less appreciation (I will come back to this later).

## **The popularity paradox**

Benefits of screening are intuitively appealing for most of us, while the harms are much harder to grasp. Laymen as well as clinicians tend to overestimate benefits of screening, while underestimating harms (Schwartz 2004, Wegwarth 2012, Hoffmann 2015). In one study of US adults, two thirds stated that they would want to be screened for cancer even if there was no treatment available, nine out of ten thought that cancer screening is always a good idea and three out of four stated that it will always save lives to find cancer early (Schwartz 2004). In a study of US physicians, three out of four stated that an improved 5-year survival rate proves that screening saves lives (Wegwarth 2012). As accounted for in the Introduction of this thesis, none of these answers are correct.

The concept of overdiagnosis is not easily conveyed to lay people (Hersch 2015). Recurrently, public media portrays celebrities who claim that their lives have been saved by screening (Chen 2013). In reality, this is impossible to know on the individual level for the majority of cases and for many people with screen-detected conditions, the risk that they have been harmed by screening may actually be greater than the chance that their lives have been saved (Welch 2011). The consequence is a popularity paradox; the more people that are (over)diagnosed through screening, the more popular screening get, no matter if the screening in question is beneficial or not (Raffle 2007).

## **Conflicts of interest in research**

The fact that researchers provide evidence for benefits but fail to do the same for harms (Jørgensen 2007, Heleno 2013, Harris 2014, De-Frank 2015) arguably contributes to the popularity of screening also

in the general public. A systematic review of randomised cancer screening trials found that the most important harm of screening, overdiagnosis, was quantified in only 7% of all screening trials (Helino 2013). Another systematic review found that the psychosocial consequences of screening were not adequately explored for any of the five screening modalities investigated (DeFrank 2015).

This is likely in part due to financial conflict of interests within research. However, even in the absence of financial gain, intellectual bias amongst researchers might still contribute to the problem. Indeed, it is common that researchers downplay or fail to report harms of the interventions they have developed and/or studied, that benefits are overestimated, and that conclusions are overstated and not supported by the data (Schwartz 1999, Glasziou 2014, Ioannidis 2014, Horton 2015). When we as researchers have dedicated time, resources and heart into a research project that we believe in, few of us are fully prepared to change our minds in accordance with the actual findings. Research careers are built on the benefits of medical interventions (or the harms of interventions, which one might suggest in my case).

Measures to tackle the problems related to intellectual and financial conflicts of interests within research have been taken. A series of papers with the overarching theme “Increasing value and reducing waste in biomedical research”, published in the *Lancet* in 2014, summarize many of these initiatives (Macleod 2014). Several recommendations were made in these publications, many of which aimed at more transparency, for example by increased use of prospectively published protocols (not only for randomised trials), sharing of raw data from studies to enable independent evaluation of the derived results, and thorough description of the methods used to enable replication of studies (Glasziou 2014, Ioannidis 2014).

But even if important progress has been made, these problems are still widespread. For example, in a recent randomised trial of screen-

ing for cardiovascular disease (abdominal aortic aneurysm, peripheral artery disease and hypertension) published in the Lancet (Lindholt 2017a), harms were very poorly reported. Additionally, a small but statistically significant reduction in total mortality made the authors conclude that "*the observed reduction of mortality risk from AAA mortality, peripheral artery disease, and hypertension*" should lead to considerations of implementing the intervention. However, the study could not demonstrate an effect on cause specific mortality of any kind. The apparent effect on total mortality was based on small numbers and was distributed between several different causes of death with no predominance of those causes targeted by the intervention (Lindholt 2017b). A possible reduction in cancer mortality contributed twice as much to the difference in total mortality as the reduction in deaths from abdominal aortic aneurysm (Lindholt 2017b). In conclusion, the exact mechanism behind the observed difference in total mortality remains elusive and the risk of random error should be considered (Jørgensen 2018). It would have been appropriate with a more cautious conclusion from the authors.

### **Conflicts of interest in policy making**

There is no common unit of measurement for benefits and harms of screening and the balance will always depend on a subjective value judgement, which no one has exclusive privilege to assess (Harris 2015, Carter 2017). Therefore, when authorities decide if a screening programme should be implemented or not, or if an existing screening programme should be deimplemented or continued, the process should be transparent and the inherent value judgements should be made explicit (Barratt 2017, Carter 2017).

Globally, the importance of having independent panels with a wide range of expertise, also from outside of medical science, to evaluate

the evidence base in medical policy making is increasingly emphasised (Barratt 2017). For example, when France recently performed an inquiry on breast cancer screening, an independent steering committee was appointed (the process of the inquiry was presented in a recent publication from Barratt and colleagues, on which the following exposition is based, Barratt 2017). The steering committee gathered health professionals with a wide range of expertise (oncology, general medicine, epidemiology, public health), as well as social sciences professionals (anthropology, law, economics, history of science, bioethics). Particular effort was taken to only include people free of both financial and intellectual conflict of interests in relation to breast cancer screening. Additionally, input from citizens with diverse socioeconomic background was carefully ensured. The inquiry recommended to either end the national breast cancer screening programme, or radically reform the current programme. This recommendation is in line with the conclusion of another independent review of breast cancer screening from Switzerland. Furthermore, in the French review, the need for complete and balanced information to potential screening participants was emphasised, as well as the acknowledgement of overdiagnosis as a serious harm. This is also in agreement with the conclusion of the Swiss review, as well as another independent review from the UK, but in contrast to many other recommendations from official bodies (Barratt 2017). Barratt and colleagues discuss what might account for the differences between the independent reviews and the other recommendations. They suggest:

”One possible explanation is that some panels may be compromised by the conflicts of interest of members, something carefully avoided in the 3 European inquiries. A broader range of disciplinary perspectives may also be important, as panel members with expertise in human and social sciences may be more likely to raise and discuss social, legal, and ethical considerations relevant to population



screening. Panels that make recommendations about medical treatments do not typically seek the values and preferences of citizens in formulating recommendations. Cancer screening programs, however, impact the lives of the public, and their preferences are important when reaching decisions.” (Barratt 2017)

It could be argued that such rigorous processes as in the French review of breast cancer screening are too expensive and resource consuming to be expected for all screening programmes in every country. I do acknowledge the dilemma for medical policy makers who might have many pressing questions to attend to. However, the costs of making rigorous and careful evaluations of screening programmes are arguably small compared to the costs of running a public screening programme (Barratt 2017).

Furthermore, processes for policy decisions today are often unacceptably far from the process described above. For example, in the Swedish Social Board of Health and Welfare’s investigation of screening for abdominal aortic aneurysm, the experts assessing the evidence base were all vascular surgeons that had been responsible or involved in implementing the screening programme in Sweden (Johansson 2016a, Johansson 2016b, Johansson 2016c). The final decision to endorse screening for abdominal aortic aneurysm was taken by a group where the vast majority were responsible for screening programmes for abdominal aortic aneurysm in their counties, and there was very limited representation of expertise outside of the community of vascular surgery (Johansson 2016a, Johansson 2016b, Johansson 2016c). Arguably, this indicates a high risk of intellectual bias, which questions the objectivity and conclusions of the investigation. We have criticised the review for being biased in favour of screening, with undue emphasis on studies suggesting a beneficial effect of screening

while ignoring evidence of harms (Johansson 2016a, Johansson 2016b, Johansson 2016c).

## Societal interventions vs screening individuals

The decision to implement, or not to implement, screening is never based on science alone (Carter 2017). Screening programmes are just as much sociological phenomena, influenced by the progression of thoughts in society as a whole, as they are a consequence of the factual numbers found in randomised trials. Western culture is permeated by a strong focus on the individual (Fugelli 2006). I wonder if our attraction towards individualistic approaches, both within medicine and in society as a whole, is a central reason to the seemingly disproportional enthusiasm for screening today?

Screening for lung cancer is estimated to save about 12 000 of the approximately 160 000 deaths from lung cancer every year in the US (Ma 2013). At the same time, lung cancer screening results in harms. Examples of such harms are overdiagnosis and subsequent major surgery with a risk of serious complications, and false positives followed by biopsies of the lungs, a procedure that is associated with medical risks (Moyer 2014). In the same country, interventions at the societal level to decrease smoking was estimated to save eight million premature deaths between 1964 to 2012, which corresponds to more than 160 000 prevented deaths in the US every year (Holford 2014), i.e. more than 10 times as many lives saved per year as by screening for lung cancer. In addition to the difference in number of lives saved, the latter intervention strategy is also cheaper, and has considerably less potential for harm. It also affects a substantial number of other diseases and cause of mortality.

Apart from societal interventions aimed directly at improving health, other societal factors aiming at social equity and flourishing, e.g. a well-functioning education system and creative child-welfare initiatives, likely by far exceed the effects of interventions within the health care system (Marmot 2012). Nobel prize winner Professor James J Heckman's research indicate that social investments in early childhood, especially for disadvantaged families, have fundamental beneficial effects on both health and economy and is not only cost-effective but actually cost saving (Heckman 2006). However, public health spending has grown rapidly in high-income countries over most of the last half a century (OECD 2015). Inevitably, increased expenditure on the health care system will take resources from other societal sectors. Policy makers have to choose where to spend the money, and currently the health care sector gets more and more of the finite resources. This development seems irrational from a public health perspective. And to spend resources on screening programmes could be questioned in this context, since even the most beneficial screening programmes have quite limited effect in absolute numbers (Prasad 2016) in addition to the previously mentioned downsides and harms.

Why do we find it much more appealing to improve public health by declaring war against life-threatening disease amongst asymptomatic individuals, rather than posing smoking bans on restaurants, or dealing with systemic causes of alienation and exclusion of vulnerable groups? Why are we so keen to cling to individual approaches to improve public health even when the effect is known to be limited, at least in relation to other possible strategies? Is this propensity to focus on the individual something deeply human? Or is it a phenomenon specific to Western culture, formed by centuries of history, the Enlightenment and the Scientific revolution?

## Medical visions and steady progress

In his thesis, Norwegian researcher Henrik Vogt discusses the role that the creation of expectations and hype play in biomedicine. He argues that in the case of genetics and precision medicine, medical researchers act as visionaries who promise continuous progression and market “simple” biomedical solutions as means to virtually extinguish ill health and suffering in the future (Vogt 2017). This is interrelated to an aspiration to control health and a reluctance to accept uncertainty in our societies. But we seem to forget that the human body is an incredibly complex system existing in an even more complex context, that causes of disease are multi-faceted, and that disease and disease progression is exceedingly difficult to foresee (Fugelli 2006, Hofmann 2015, Vogt 2017).

Inflated and irresponsible visions of the effects of medical interventions might redirect resources within the health care system, or from societal interventions to the health care system, which might in turn have ruinous effects for the health and well-being of the population (Pétursson 2012). In her thesis, Icelandic researcher Margrét Ólafía Tómasdóttir, showed that nearly half of the adult Norwegian population (20 to 79 years) had multimorbidity according to the classical definition (two or more chronic conditions), and nearly one fourth when defined as three or more chronic conditions. Considering that Norway has one of the most long-living and healthy populations in the world, this epidemic of medical diagnoses arguably signals a questionable expansion of the territory of medicine (Tómasdóttir 2016). Another Icelandic researcher, Hálfván Pétursson, showed in his thesis that if evidence-based guidelines for cardiovascular prevention were fully applied in Norway, more general practitioners than existed at the time of the study would be required to handle hypertension only, i.e. there would be no time for general practitioners to take care of ill patients (Pétursson 2012).

I argue that mass screening fits well into a narrative evolving around medical visions and steady progress. Researchers and policy makers have great influence on the future path of medicine. Instead of providing attractive but partially unrealistic visions, we ought to shoulder this responsibility in a reflective, sensible and conscious way (Getz 2006, Fugelli 2006, Vogt 2017).

## Consequences of the evidence hierarchy

An interview study of women with osteoporosis detected at a bone-scan revealed that the asymptomatic diagnosis fostered a new body image. The women interpreted the scan result to mean bodily fragility, which they incorporated into their bodily perception; they started to experience themselves as weak with reduced capacity (Reventlow 2006).

Such existentially loaded harms of screening are not easily captured by the research methods at the top of the evidence hierarchy of evidence-based medicine (Munn 2014, Shaw 2014, Greenhalgh 2015). Indeed, frontrunners of evidence-based medicine have long acknowledged that different research questions require different methods; the perception that randomised trials always present the best answer is thereby flawed. Questions about prognosis, aetiology, diagnosis, disease frequency and adverse events are often best answered with other study designs (Glasziou 2004). Arguably, carefully designed qualitative studies are needed to explore and fully appreciate the subtler and more existential harms of screening. However, knowledge from qualitative studies is generally downplayed in evidence synthesis and policy making (Munn 2014, Shaw 2014, Greenhalgh 2015). This points to a fundamental problem within evidence-based medicine and medical policy making today, namely that the evidence hierarchy tends to di-

rect our focus towards outcomes that are easy to measure, at the expense of the more indistinct outcomes that are difficult to measure (Greenhalgh 2015, Kelly 2015). Referring again to my light post metaphor, crucial aspects of medical interventions end up between the light posts and are not picked up. Because of the way evidence-based medicine is (mal)practiced today, it does not only become a tool to inform us about the available evidence for a certain intervention, but also forms our perception of what is important (Forssén 2011).

## Screening – deeper and broader meaning

Inherent in medical science today lies an ideal of value-neutrality (Kelly 2015, Carter 2017), described below by Forssén and colleagues (Forssén 2011):

“A common metaphor for science is a cone of light illuminating an increasing part of a dark map. In this image of science.. ..the scientist has no personal location, and is invisible, signifying objectivity and scientific freedom.. ..The most important consequence of this ‘functional myth’ is that the scientific community cannot be held responsible for what it simply ‘uncovers’. Therefore, the scientist has no moral responsibility.”

An often-unquestioned framework of medical interpretation and classification permeates research (Lock 1988, Getz 2006, Forssén 2011). But knowledge must be interpreted in relation to the context in which it is created (Haraway 1991). By directing “the cone of the light” researchers, and funders of research, have great influence on society’s understanding of what is counted as disease, and what is considered as cause, prevention and cure (Forssén 2011).

As a consequence of the increased focus on prevention and early detection within Western medicine, many diagnoses today do not exist in their own right, but might be interpreted as artefacts created by increasing medical (sub)classification and (over)investigation (Getz 2006, Pétursson 2012, Tómasdóttir 2017). Asymptomatic diseases and risk factors found at screening are examples of this (Martin 2014, Järvinen 2015), as well as a myriad of medically unexplained symptoms. Other examples involve expanding disease definitions (Moynihan 2013, Cundy 2014), and the fact that more and more of ordinary life circumstances are defined as medical problems (Dowrick 2013, Meixel 2015).

This development is underpinned by values. For example; randomised trials cannot tell us whether symptoms of stressful life circumstances should be labelled as depression or not, or if we should give the majority of the adult population pills for cardiovascular risk factors. This is because these decisions are at least partly based on normative, not empirical, questions. No amount of data can thereby justify them. But the underpinning values are seldom reflected upon in the scientific publications. And evidence syntheses' rarely zoom out to reflect on the construct of the disease or the intervention in question, or the broader consequences for society as a whole. In this way even the good systematic reviews and evidence-based guidelines, with adequate evaluation of the evidence base and no vested interests, may contribute to establish a certain understanding of symptoms, medical risk or interventions, when it might have been a more sensible approach to instead question the underlying constructs themselves (Forssén 2011).

Departing from the works of Ivan Illich (Illich 1974), in the context of screening for non-communicable life-threatening disease, questions that in my opinion have gained too little attention include:

How can the concept of approaching asymptomatic citizens and informing them that they might have a dangerous disease be understood from a value-perspective? Does screening affect our perceptions of health in society as a whole? Does the diagnosing of asymptomatic people with life-threatening disease influence our ability to trust that our bodies will tell us if we are sick and in need of help? Does screening change our attitude from assuming that we are healthy until proven otherwise, to assuming that we are sick until proven otherwise? And in that case, can this conception affect our health and well-being? What impact do screening programmes have on our ability to tolerate ill health, disease and even bad luck, in society as a whole? Does the increased focus on early detection influence our capability to deal with the inevitable uncertainty associated with living? How does screening various body parts of our patients fit in with person-centred care and “seeing the whole person” - as opposed to fragmented and reductionist health care? What does the increased focus on early detection and screening mean for the future direction of medicine?

Such questions are not easily answered by science alone; high quality evidence from randomised trials will not help us deal with these issues. Although qualitative studies could contribute with a deeper understanding, we also need to reflect on dimensions other than the strictly scientific. Researchers need to contemplate on the influence that their research will have on society as a whole in a much broader sense (Getz 2006, Fugelli 2006, Forssén 2011), and policy makers need to do likewise in regards to their decisions about health care. We should take a step back, and view our own ideas within the landscape of a wider history of ideas.



## The ethical dilemma of screening

The fundamental ethical problem of screening programmes is that success comes at a cost; in order to improve the prognosis for some people, others will be exposed to unintended harm (Shickle 1994). There is no scientifically correct answer to whether it is ethically acceptable for the health care system to cause harm to asymptomatic people through an intervention they have not themselves asked for. From a utilitarian perspective, screening is justifiable if it results in net benefit (Shickle 1994, Kelly 2015). This is of course difficult to judge when adequate evaluation of harms is lacking for most screening programmes. Moreover, even if harms were adequately explored and investigated, science cannot provide an answer on what constitutes a net benefit; to weigh benefits and harms against each other is to compare apples and oranges (Shickle 1994, Harris 2015). How many people can undergo unnecessary surgery of an abdominal aortic aneurysm in order to justify the prevention of one death from the disease? How much distress from being labelled with a diagnosis of malignant melanoma, to how many people, is acceptable in order to prevent one melanoma-death? These decisions are value judgements, and need to be recognised as such (Harris 2015, Kelly 2015, Carter 2017).

Moreover, strict utilitarianism is hardly acceptable in our societies. We must also consider overarching deontological principles, such as “first of all, do no harm” (*primum non nocere*) (Shickle 1994). This principle is of course not entirely applicable to medicine in general, since almost all interventions have harms, but may nonetheless be highly justified from an ethical perspective. For example, in people with aggressive cancer, chemotherapy with high curative rates can arguably be ethically justifiable, even if this treatment has terrible side effects. The crucial difference between treatment of symptomatic cancer and screening is that in screening, harms are inflicted to people who would never have had any symptoms if they had not participated in

the intervention. When inviting asymptomatic citizens to an intervention they have not asked for, a great responsibility follows (Sackett 2002). In my opinion, this responsibility is seldom shouldered today, and health authorities seem to make light of the ethical dilemmas that screening programmes entail.

## Informed choices – a solution?

Informed choice and shared decision making are increasingly proposed as means to mitigate the ethical dilemmas presented by screening (Brownsword 2010, Woloshin 2012, Hoffmann 2014). Personal preference is emphasised in modern medicine and putting the responsibility on the shoulders of the individual seems in line with patient-centred care. At the same time, a paternalistic discourse is apparent in screening: an overwhelming amount of studies are done with the explicit agenda to increase uptake in screening (Everett 2011), without consideration of whether participation is based on informed choice or not (Ploug 2012). A shift of perspective, from paternalism to respect for people's autonomy, is one important step towards an approach consistent with contemporary ethical values (Ploug 2012).

However, many screening programmes do not provide information about harms in invitations today (Jørgensen 2009, Gøtzsche 2011, Ploug 2012, Kolthoff 2016), which overrides the autonomy of the individual. Further, many people will probably trust that an invitation from health authorities strongly indicates that the intervention is worthwhile. This assumption is emphasised by widespread use of pre-booked appointments (versus opt-in alternatives), which short-circuit the decision process by indicating a seemingly correct and expected choice (Ploug 2012). Consequently, there is a need for a reassessment of invitations to current screening programmes; adequate information

on all important benefits and harms should be added, in addition to an acknowledgment that non-participation might be as rational a choice as participation.

Balanced comprehensive information is important from an ethical perspective; however, it might not have a substantial effect on the ability of people to make truly informed choices based on truly personal preferences. Inherent in the idea of informed choice about screening participation lies an assumption that information speaking to our intellect is easily integrated into our understanding of risk. Yet research suggests that our understanding of risk relies mainly on emotions and that cognitive comprehension has little effect on decision-making (Loewenstein 2001, Slovic 2002). If emotionally charged messages have formed our perception of a particular intervention, which is certainly the case for many types of screening, subsequent information is unlikely to change our understanding of that intervention nor our attitudes or behaviour (Slovic 2002).

Furthermore, in the concept of informed choice lies an understanding about “personal preferences” that could be questioned, as done in a paper by Nelson and colleagues (Nelson 2007):

“it is not clear that people have pre-existing preferences and values that merely need to be revealed through some elicitation process. In many situations it makes little sense to speak of “clarifying” preferences because preferences may be created de novo in response to the values elicitation process.”

The concept of personal preferences in regards to screening becomes even more complex when considering that “taking care of ones health” (for example by participating in screening, or checking ones blood pressure or blood lipids) is considered more or less a societal responsibility today, an attitude underpinned by vested interests with-

in the “health industry” (Fugelli 2006, Ploug 2012). Indeed, it is irresponsible to consider patient preferences without taking into account that strong market forces are pushing peoples preferences in the direction of more medicine.

Personalised risk communication has been proposed as a means to increase informed choice about screening (Edwards 2013). Although appealing, are such methods sound from a public-health perspective? Should we use scarce resources to maximise informed choice among healthy individuals when money could instead be spent on people with the greatest need, those who are already ill?

In conclusion, balanced information in invitations to screening seems merely a small step towards true informed choices based on truly personal preferences. Equally important is a better understanding of the benefits and harms of screening among journalists, clinicians, researchers, editors, and policy makers and a more nuanced debate about screening in the media (related paper 4 in Appendix).

However, even if truly informed choices based on truly personal preferences were possible, it would still be ethically problematic to approach healthy people and inform them that they might have a dangerous disease and that screening could save them. At the same time there are serious harms that might outweigh the benefit, which the invitee must also consider. On the basis of this information, they must now take responsibility and make an informed choice about participation. This responsibility for the decision cannot be avoided, because the invitation itself makes a choice mandatory – we have placed the individual in a situation with no possibility to avoid making a choice, and in the process we have transferred responsibility from the health care provider to the individual (related paper 1 in Appendix).

Who considers the ethics and takes responsibility for presenting people with such a complex choice in the first place? When discussing

the ethics of informed choice, it is essential to consider the locus of initiative. That is, who started the process leading to the need for an informed choice? The ethics of informed choice is fairly uncomplicated in situations where clear-cut needs of the patient are the driver for an unambiguous diagnosis. But screening is not driven primarily by the patient's own agenda, which changes the ethical premises of informed choice (related paper 1 in Appendix).

Informed choice must not be used to justify the introduction or continued use of screening programmes in which the balance between benefits and harms is doubtful, and informed choice does not remove the responsibility for offering such screening programmes from health authorities. A strong focus on informed choice and shared decision making in preventive medicine entail ethical considerations that have gained too little attention. Furthermore, true informed choice remains utopic in our cultural context and does not solve the fundamental dilemma of screening; is it ethically acceptable to cause serious harm in some people to improve the prognosis of others?



# Conclusion

## Papers I and II

Screening for malignant melanoma is widely adopted in the Western world although there is no data from randomised trials on the benefits and harms of the intervention. Due to uncertainty of benefit, and risk of harm through overdiagnosis and opportunity costs, opportunistic or organised screening for malignant melanoma should not be implemented or recommended outside the confines of a well-designed, randomised trial.

## Paper III

Screening for abdominal aortic aneurysm was introduced in several countries without adequate exploration and quantification of harms. Likewise, a lowered cut-off for the definition on an abdominal aortic aneurysm within the screening programmes is at the brink of implementation without adequate consideration of potential harms.

## Paper IV

Screening for abdominal aortic aneurysm had only a minor effect on disease-specific mortality in Sweden; only 7% of the benefit estimated in the largest trial of screening for abdominal aortic aneurysm was observed in our study. Large reductions in mortality from abdominal aortic aneurysm were present in both screened and non-screened co-

horts and thus mainly caused by other factors, likely reduced smoking. The clinical significance of screening for abdominal aortic aneurysm in contemporary populations is therefore questionable. The absolute number of overdiagnosed and overtreated cases caused by screening for abdominal aortic aneurysm was also reduced compared to the randomised trials, but not to the same extent (28% and 51% respectively). The small benefit and substantially worsened benefit-to-harm balance questions the continued justification of screening for abdominal aortic aneurysm.

## General conclusion

The overarching conclusion of this thesis is that benefits of screening receive much more attention and appreciation than harms, both within research and in policymaking. This is problematic from an ethical perspective, since inviting asymptomatic citizens to an intervention they have not asked for entails a great responsibility not to cause more harm than good. Further, the cases of screening for malignant melanoma and screening for abdominal aortic aneurysm points to some of the fundamental problems inherent in medical thinking, science and policy-making today. Namely; a disregard of downsides and harms of interventions, a negligence towards appropriate research methods for investigating “psychosocial” harms of medical interventions, an inability to manage intellectual bias in research as well as in policy making, and an inability to consider inappropriate resource allocation and opportunity costs of medical interventions in general, and preventive medicine in particular. Furthermore, we fail to critically reflect on the fundamental values behind the very idea to screen asymptomatic citizens for early stages of disease as a means to improve public health.



# Future perspective

Thanks to the work of brilliant people, considerable progress in regards to the problems outlined in the General Discussion of this thesis has been made. Increased attention is given to harms of screening (Harris 2014, Barratt 2017). For example; in the last decades the concept of overdiagnosis has gone from a “shadowy idea to acknowledged reality” (Barratt 2015). Measures are taken to counteract problems due to financial and intellectual conflicts of interests in research, for example through increased transparency by registration of study protocols and data sharing (Glasziou 2014, Ioannidis 2014). The importance of independent committees with a broad representation of diverse competences when it comes to policy decisions within health care is increasingly acknowledged (Barratt 2017). Patient involvement in research, evidence synthesis and guideline development, to ensure that outcomes relevant to patients are given due consideration, is getting more common (Chalmers 2014). The importance of patient preferences is increasingly emphasised in clinical practice, for example through shared decision-making processes (Hoffmann 2014).

However, it could be argued that the problems outlined in the General Discussion of this thesis are already out-dated. “Precision medicine” is on the verge to medicalizing almost every aspect of human life through it-words like genomics, proteomics and metabolomics. Whole-body scans and extensive gene testing is heavily marketed under the pretence of enhancing personal freedom (that is for the minority of the population who can afford it). Conventional screening programmes for single diseases seem yesterdays matter. Likewise with

the mandate for health authorities to decide which screening that should be implemented, and which should not. With the technology of tomorrow, our state of health can easily be picked up (and also understood?) by a billion data-points, which could be read in real time from apps on our mobile phones. Lack of evidence for any benefit of such mass-scale entrance of “big data” into Western medicine does not seem to be an issue. In terms of harms like overdiagnosis and overtreatment I would like to quote the Norwegian researcher Henrik Vogt; “We ain’t seen nothing yet”.

Arguably, the existential, philosophical and humanistic aspects of medicine are disregarded within our scientific paradigm. The downstream effect is a fragmented and reductionist understanding of health and well-being (Kirkengen 2016). “Precision medicine” risks adding enormous amount of fuel to this development. But I have hope. The pendulum inevitably swings.

In my opinion, Western medicine needs to remove the mask of presumed scientific neutrality and acknowledge the profound influence of values, not just in regards to patient preferences but in every step of the process of research and policy making (Kelly 2015, Carter 2017). We should (and I believe we will) critically reflect on the deeper meaning and the broader consequences of what we do in medicine to a much larger extent, and in a much more structured way, than what we do today. As a reference to the picture on the cover of this thesis, we should go outside of the territory that the streetlights illuminate, equip ourselves with torches and approach the wild and rocky terrain that hides in the dark. Humbly keeping in mind that where we choose to direct our cone of light will decide what we will see.



Photographer: Sebastian Dijkstra



# Epilogue

As I have argued in this thesis, I believe that it is fundamentally important to explore and critically reflect on the inherent values of our research. I will now reflect on the values underpinning my own research.

I am hesitant to the idea of screening asymptomatic people for disease or risk factors. I do think that we need to be much more careful when we actively interfere with the lives of asymptomatic people than we are today. I believe that preventive medicine is good, but only when it is confined to the interventions where the benefits clearly outweigh the harms, the ethical premises are deemed acceptable after careful consideration, and the consequential resource allocation is adequately explored and sensible. In my opinion, this does not apply to many of the preventive interventions within Western medicine today. And it seems to me like we are heading towards accepting exceedingly less favourable benefit-to-harm balance when it comes to preventive medicine, which further exacerbates a resource allocation from those with the greatest need; the already ill.

Much of these viewpoints are probably rooted in my experiences from working as a general practitioner. In my daily work, I feel a frustration due to what I regard as an unsound and unfair use of my time. I spend many hours every day on preventive medicine for people with low risk of developing symptomatic disease. At the same time, I do not have enough time to adequately take care of those who need me the most; for example people with symptoms that could indicate serious disease and therefore are in need of swift work-up, or people

whose suffering might be soothed by a doctor who cares and takes time to listen.

Furthermore, the majority of patients for whom I provide preventive medicine are people with rather high socioeconomic status. People with low socioeconomic status, or people alienated from society, do not seem to show up for preventive medicine at my clinic. I find this hugely problematic, because it augments inequity in society. Ironically, low socioeconomic status probably confers a higher risk for many diseases, than the risk factors that I measure and treat at my health care centre. Therefore, I find the social gradient in the utilization of preventive medicine not only unethical, but also foolish.

Another preconception that I believe influences my research is my scepticism towards the prevailing reductionism within Western medicine. We tend to think about our patient's symptoms in terms of diagnoses, body parts or organs, and too often fail to see the whole person. I believe that inherent in medicine today, lies an unquestioned framework of medical interpretation and classification that directs our mind paths towards a simplified understanding of complex phenomena. Our biomedical framework for understanding disease makes us sort our patients' illnesses and suffering into diagnoses that may be technically correct but not necessarily existentially meaningful in the sense of enhancing the patients' ability to engage in life.

Further, my natural reference when it comes to screening for abdominal aortic aneurysm is a patient that I have known for many years, who after being diagnosed with an aneurysm detected at screening does not dare to have sex anymore out of fear of rupture, and who suddenly starts to seek me repeatedly due to worries about multiple conditions (unrelated to the screening diagnosis) although being free of symptoms. Or a patient undergoing elective surgery for an aneurysm, with "no complications" stated in the records from the hospital, but yet not being able to return to his previous independent

life, but having to move to a nursing home. I imagine that for many vascular surgeons, the natural reference in regards to screening for abdominal aortic aneurysm is a man dying in the emergency room from a ruptured aneurysm, a death that might have been possible to avoid if only the aneurysm would have been detected in time. None of these references are more valid than the other. But I believe that the perspective we have will inevitably influence our research. I believe that my perspective has influenced the choice of the research questions of this thesis, the methods used, and the interpretation of the results.

One aspect that I would have done differently if I had done this thesis all over again, is involving someone with a very different perspective than mine in the projects. Arguably, my co-authors have more or less the same perspective as me when it comes to screening. I did try to recruit a vascular surgeon to Paper IV of this thesis, but I was unable to find someone willing to take part. I sympathise with this reluctance. Because of the polarized environment on this issue, I would probably be hesitant to take part in research projects run by the community of vascular surgeons.

I hope that I have hereby offered transparency about my preconceptions and inherent values. I encourage readers of this thesis to critically reflect on the research findings and the arguments presented in this thesis, in the light of the values underpinning them.





# Acknowledgement

I want to start with thanking all the male chauvinists who I have encountered in the process of writing this thesis. You did get under my skin. But eventually, the shame transformed into anger. An anger that fuelled my energy and gave me the strength to keep on. I could not have done this without you.

Much more importantly, I am extremely grateful to all the awesome people who have supported, inspired and challenged me, and I feel a humble gratitude to all the brilliant forerunners of this research field, on whose ideas this thesis rests.

I want to express my great appreciation to the most patient man I know; Bertil Marklund, my main supervisor. Thank you for tirelessly helping me through the endless and sometimes very challenging hazards that writing this thesis brought to us. To have you as a supervisor has been rewarding, educative, comforting, and fun! You have done so much of the hard work without getting the credit for it. Luckily, you are a man with great wisdom, and a very small ego.

I want to thank my co-supervisor Ander Hansson for opening my eyes to the fact that screening for abdominal aortic aneurysm is not only an extremely interesting topic, but also a primary care research topic. It had profound importance for my research path, which I am very grateful for today. I want to thank my other co-supervisor John Brodersen, especially for so generously inviting me to the research environment at the university of Copenhagen and letting me ride the wave of your impressive international research network. The effort

you put in helping new researchers to build networks and platforms is indeed very inspiring.

Thank you Karsten Juhl Jørgensen for always providing thorough and critical feedback (returning everything I ever send to you full of **red marks**). I have learned so much from you. Moreover, you have always taken the time to support me. I especially remember one occasion; I had made a fool out of myself and it did not take long before I got a long and terrifying mail from an opponent. You immediately called me, laughed it all off and said “You should be proud to have caused such strong feelings, it means you’re on to something important”. So instead of dropping out of my thesis, I took a snus and rolled up my sleeves.

I want to thank Volkert Siersma and Per Henrik Zahl, statisticians on Study IV of this thesis, for their tireless, hard and progressive work on the project. I have learned a lot from the two of you!

I want to acknowledge Bertil Marklund, Hálfván Pétursson, Linn Getz, Karsten Juhl Jørgensen, Peder Halvorsen and Katarina Hamberg for providing intellectually stimulating feedback on the frame of this thesis, thank you for your generosity!

I warm thank you to my “older” siblings in research, Hálfván Pétursson and Margrét Ólafía Tómasdóttir. It is such a relief to rest in your company. Thanks for everything from sharing your intelligent ideas, to comforting ego-boost mails, to cheering me up with your dark and sharp humour, to care for me when I am not quite able to keep up with the Icelandic drinking habits. You are simply brilliant!

I want to thank Linn Getz and Jóhann Ágúst Sigurdsson. It was a mind-blowing lecture from Linn in 2010 that evoke my interest in research. After that lecture, I wrote a mail to my idol, and she answered with impressive commitment. Thank you for welcoming me into your home, for the extraordinary breakfasts, organising of show-

er schemes, and for delicious cakes and caring notes on the table when your “academic children” got home late at night. Your efforts are not just enjoyable; they are the creation of a flourishing research environment.

I want to thank Peder Halvorsen for your considerate encouragement, and Anna Luise Kirkengen for always taking the time to engage in my troubles. Thanks also to the rest of the Nordic Risk Group, you have been a great support and inspiration to me! A special thanks to Birgitta Hovellius, Annika Forssén, Susanne Reventlow, Göran Sjönell, Eivind Meland, Josabeth Hultberg, Lotte Hvas, Katarina Hamberg, Jan Håkansson, Bente Prytz Mjølstad, Hans Blomberg, Lars Englund, Andreas Pahle and Henrik Vogt. And to Irene Hetlevik for starting the group many years ago, your effort has indeed had extensive effects.

I want to thank the extraordinary Gemma Jacklyn from the University of Sydney (aka “the queen of networking”), whom I greatly admire, for letting the social misfits from the Nord ride on your networking skills. Your contagious laughter could make a stone smile!

I want to thank Matteo Bruschetti, director of Cochrane Sweden, for being so generous and collaborative. I am looking forward to long and fruitful discussions about feminism with you in the near future. Be aware!

A huge thanks to all the staff at Herrestads Health Care Centre! Thank you for never complaining about my absence and always trying to facilitate my research. A special thanks to my fantastic colleagues Kerstin Holmberg, Inger Grönlund Brown, Lars Dotevall, Karin Mossberg, Jenny Skälegård and Ellinor Fredriksson, for always being there for me in times of need. To face the hardships of clinical reality together with you is emotionally rewarding and intellectually stimulating. A special special thanks to Karin Mossberg – it is such a privilege

to be able to share research experiences with you on a daily basis. Another special thanks to my boss Pia Zuhr and to my unofficial boss Maria Blidberg - for all the times you gave me time off with short notice (and for always managing to do it with a smile). Also a warm thank you to my former boss Per Kjellson, for your incredible support in the beginning of my PhD.

I want to thank the local research unit (FoUU Fyrbodal), for your support and for funding part of my PhD. A special thanks to Sven Kylén, Eva Larsson and Anna Bergenheim.

I am very grateful to the National Research School for Primary Health Care. The research network that the school has created has had profound importance for me. A special thanks to Lars Hjalmar Lindholm, Maria Boström, Cecilia Björkelund, Kristina Boström, the PhD students in my group, Simon Griffin, and to Stuart Spencer - for always challenging every single aspect of every single one of our ideas.

I want to thank Iona Heath, for being so tremendously inspiring – every time I get filled with doubt about the meaning and purpose of my work, reading a randomly chosen text from Iona keeps me on track.

I want to thank Staffan Björck and Samuel Adamsson Eryd from the Centre of Registers Västra Götaland, for very valuable advice on how to get by in the jungle of registry based research.

I want to thank Mie Sara Hestbach from the University of Copenhagen, Bruno Heleno from the University of Lisbon, Jolyn Hersch from the University of Sydney, and Hazel Thornton. You were a great support to me in the beginning of my PhD.

I want to thank my dear friends Karin Mossberg, Karin Rydenman and Jenny Skälegård for tolerating my asocial personality and for your intelligent input to my research ideas (as well as for enduring listening

to them for a quite extensive amount of time). See you on the dance floor!

A warm thanks to my dear old friend and the only (but very enthusiastic) member of my fan club, Helena Gotte. I greatly admire your independence and strength, and thanks for not giving up on me!

I want to thank my parents in law Ingalil, Kenth and Helene, for taking care of the kids on numerous occasions when my research collided with family life – and for always supporting and encouraging me.

I want to thank my mother Eva and my father Peter. You have, both in your own ways, always encouraged me to think. Even when I was very little, you elucidated the mysteries of the world in a serious mode, taking for granted that I would understand also the most complicated issues of our inner and outer universes. The seeds to many of my research ideas spring from discussions I have had with the two of you. Also, you have instilled in me the importance of solidarity, and for that I am grateful.

A warm and loving thanks to my brilliant sisters Hanna, Lina, Hulda and Karna. You are, and always will be, the fundament from which I dare to reach out to the world. Without you, I would be lost.

Thank you Iris and Sven, for being you! You bring perspective and joy to my life. And in regards to this thesis, you have been absolutely fantastic. Every time I published a paper, you made me (at least) one delicious cake, and it always tasted like a little bit of heaven.

Fredrik, there is so much I could write about you. I admire and respect you on so many levels. But I know that you will hate to read this. So for your sake, I will try to keep most of the sentimentality to myself. When the conflicts related to my research was tearing me down and making me a terrible mother and an even worse wife, I told

you that I had decided to drop out of my PhD - I considered the cost to my family too high. You looked at me with love in your eyes and said; “No Minna, you need this. Without research, you would be a bitch”. And you gave me the time and space that I needed. Needless to say, I would never ever have been able to do this without your extraordinary support. Even in one of the most gender equal countries of the world, not many women have the opportunities that you have given me.

# References

1. Ahn HS, Kim HJ, Welch HG. Korea's thyroid-cancer "epidemic" – screening and overdiagnosis. *N Engl J Med* 2014;371:1765-7.
2. American Cancer Society. Skin Cancer Prevention and Early Detection. Accessed 5 January 2017. Available at: [www.cancer.org/cancer/cancercauses/sunanduvexposure/skincancerpreventionandearlydetection/skin-cancer-prevention-and-early-detection-skin-exams](http://www.cancer.org/cancer/cancercauses/sunanduvexposure/skincancerpreventionandearlydetection/skin-cancer-prevention-and-early-detection-skin-exams)
3. American Academy of dermatology. 30 years of free skin cancer screening. Accessed 5 January 2017. Available at: [www.aad.org/public/spot-skin-cancer/programs/screenings/30-years-of-skin-cancer-awareness](http://www.aad.org/public/spot-skin-cancer/programs/screenings/30-years-of-skin-cancer-awareness)
4. Andermann A, Blancquaet I, Beauchamp S, Déry V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bulletin of the World Health Organization* 2008;86:317-9.
5. Ashton HA, Gao L, Kim LG, Druce PS, Thompson SG, Scott RA. Fifteen-year follow-up of a randomized clinical trial of ultrasonographic screening for abdominal aortic aneurysms. *Br J Surg* 2007;94:696-701.
6. Australian Cancer Network Melanoma Guidelines Revision Working Party 2008. Clinical Practice Guidelines for the Management of Melanoma in Australia and New Zealand. Accessed 14 June 2016. Available at: [www.cancer.org.au/content/pdf/HealthProfessionals/ClinicalGuidelines/ClinicalPracticeGuidelines-ManagementofMelanoma.pdf](http://www.cancer.org.au/content/pdf/HealthProfessionals/ClinicalGuidelines/ClinicalPracticeGuidelines-ManagementofMelanoma.pdf)
7. Balanda KP, Lowe JB, Stanton WR, Gillespie AM. Enhancing the early detection of melanoma within current guidelines. *Austr J Public Health* 1994;18:420-3.
8. Barratt A. Overdiagnosis in mammography screening: a 45 year journey from shadowy idea to acknowledged reality. *BMJ* 2015;350:h867.
9. Barratt A, Jørgensen KJ, Autier P. Reform of the National Screening Mammography Program in France. *JAMA Intern Med* 2018;178:177-8.
10. Basnyat PS, Biffin AH, Moseley LG, Hedges AR, Lewis MH. Mortality from ruptured aortic aneurysm in Wales. *Br J Surg* 1999;86:765-70.
11. Bengtsson H, Sonesson B, Bergqvist D. Incidence and prevalence of abdominal aortic aneurysms, estimated by necropsy studies and population screening by ultrasound. *Ann NY Acad Sci* 1996;800:1-24.
12. Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer overdiagnosis with mammography screening: a systematic review. *Lancet Oncol* 2007;8:1129-38.
13. Black WC, Haggstrom DA, Welch HG. All-cause mortality in randomized trials of cancer screening. *J Natl Canc Inst* 2002;94:167-73.
14. Borland R, Meehan JW. Skin examination for signs of cancer. *Austr J Public Health* 1995;19:85-8.
15. Breslow A. Thickness, cross-sectional areas and depth of invasion in the prognosis of cutaneous melanoma. *Ann Surg* 1970;172:902-8.
16. Bretthauer M, Kalager M. Principles, effectiveness and caveats in screening for cancer. *Br J Surg* 2013;100:55-65.
17. Brodersen J, McKenna SP, Doward LC, Thorsen H. Measuring the psychosocial consequences of screening. *Health Qual Life Outcomes* 2007;5:3.

18. Brodersen J, Siersma V, Ryle M. Breast cancer screening: "reassuring" the worried well?. *Scand J Public Health* 2011;39:326-32.
19. Brodersen J, Siersma VD. Long-term psychosocial consequences of false-positive screening mammography. *Ann Fam Med* 2013;11:106-15.
20. Brown P, Zelt D, Sobolev B. The risk of rupture in untreated aneurysms: the impact of size, gender, and expansion rate. *J Vasc Surg* 2003;37:280-4.
21. Brownsword R, Earnshaw JJ. The ethics of screening for abdominal aortic aneurysm in men. *J Med Ethics* 2010;36:827-30.
22. Calonge N. Screening for abdominal aortic aneurysm, recommendation statement. US preventive services taskforce 2005. [www.uspreventiveservicestaskforce.org/uspstf05/aaascr/aaars.pdf](http://www.uspreventiveservicestaskforce.org/uspstf05/aaascr/aaars.pdf).
23. Canadian Task Force on Preventive Health Care 2013. Melanoma Critical Appraisal Report. Accessed 14 June 2016. Available at: [canadiantaskforce.ca/appraised-guidelines/2013-melanoma/](http://canadiantaskforce.ca/appraised-guidelines/2013-melanoma/)
24. Carmichael K, Nolan SJ, Weston J, Tudur Smith C, Marson AG. Assessment of the quality of harms reporting in non-randomised studies and randomised controlled studies of topiramate for the treatment of epilepsy using CONSORT criteria. *Epilepsy Res* 2015;114:106-13.
25. Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *BMJ* 2015;350:g7773.
26. Carter SM. Valuing health care improvement: implicit norms, explicit normativity, and human agency. *Health Care Anal* 2017; (Epub ahead of print).
27. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gulmezoglu AM, Howells DW, Ioannidis JP, Oliver S. How to increase value and reduce waste when research priorities are set. *Lancet* 2014;383:156-65.
28. Chen JY, Eborall H, Armstrong N. Stakeholders' positions in the breast screening debate, and media coverage of the debate: a qualitative study. *Critical Public Health* 2013;19:1-11.
29. Choke E, Vijaynagar B, Thopson J, Nasim A, Bown MJ, Sayers RD. Changing epidemiology of abdominal aortic aneurysms in England and Wales: older and more benign? *Circulation* 2012;125:1617-25.
30. Copp T, Jansen J, Doust J, Mol BWJ, Dokras A, McCaffery K. Are expanding disease definitions unnecessarily labelling women with polycystic ovary syndrome?. *BMJ* 2017;358:j3694.
31. Cosford PA, Leng GC, Thomas J. Screening for abdominal aortic aneurysm. *Cochrane Database Syst Rev* 2007;2:CD002945.
32. Cundy T, Ackermann E, Ryan EA. Gestational diabetes: new criteria may triple the prevalence but effect on outcomes is unclear. *BMJ* 2014;348:g1567.
33. Damiano R, Lorenzo GD, Cantiello F, De Sio M, Perdonà S, D'Armiento M, Autorino R. Clinicopathologic features of prostate adenocarcinoma incidentally discovered at the time of radical cystectomy: an evidence-based analysis. *Eur Urol* 2007;52:648-57.
34. Darwood R, Earnshaw JJ, Turton G, Shaw E, Whyman M, Poskitt K, Rodd C, Heather B. Twenty-year review of abdominal aortic aneurysm screening in men in the county of Gloucestershire, United Kingdom. *J Vasc Surg* 2012;56:8-14.
35. Davis M, Harris M, Earnshaw JJ. Implementation of the National Health Service abdominal aortic aneurysm screening programme in England. *J Vasc Surg* 2013;57:1440-5.
36. DeFrank JT, Barclay C, Sheridan S, Brewer NT, Gilliam M, Moon AM, Rearick W, Ziemer C, Harris R. The psychological harms of screening: the evidence we have versus the evidence we need. *J Gen Intern Med* 2015;30:242-8.
37. Dinnes J, Matin RN, Moreau JF, Patel L, Chan SA, Chuchu N, Bayliss SE, Grainge M, Takwoingi Y, Davenport C, Walter FM, Fleming C, Schofield J, Shroff N, Godfrey K, O'Sullivan C, Deeks JJ, Williams HC. Tests to assist in the diagnosis of cutaneous melanoma in adults: a generic protocol [Protocol]. *Cochrane Database Syst Rev* 2015;10:CD011902.
38. Dowrick C, Frances A. Medicalising unhappiness: new classification of depression risks more patients being put on drug treatment from which they will not benefit. *BMJ* 2013;347:f7140.
39. Edwards AGK, Gurudutt N, Ahmed H, Elwyn GJ, Pickles T, Hood K, Playle R. Personalised risk communication for informed decision making about taking screening tests. *Cochrane Database Syst Rev* 2013;28:CD001865.



40. Etzioni R, Gulati R. Recognizing the limitations of cancer overdiagnosis studies: a first step towards overcoming them. *J Natl Cancer Inst* 2015;108:djv345.
41. European Association of Dermato Oncology. Euro Melanoma. Accessed 14 June 2016. Available at: [www.euromelanoma.org](http://www.euromelanoma.org)
42. Everett T, Bryant A, Griffin MF, Martin-Hirsch PL, Forbes CA, Jepson RG. Interventions targeted at women to encourage the uptake of cervical screening. *Cochrane Database Syst Rev* 2011;5:CD002834.
43. Federman DG, Concato J, Caralis PV, Hunkele GE, Kirsner RS. Screening for skin cancer in primary care settings. *Archives of Dermatological Research* 1997;133:1423-5.
44. Federman DG, Kravetz JD, Haskell SG, Ma F, Kirsner RS. Full-body skin examinations and the female veteran: prevalence and perspective. *Archives of Dermatological Research* 2006;142:312-6.
45. Ford JS, Ostroff JS, Hay JL, Buckley TR, Stein TR, Berwick M, Primavera LH, Shike M. Participation in annual skin cancer screening among women seeking routine mammography. *Preventive Medicine* 2004;38:704-12.
46. Forssén A, Meland E, Hetlevik I, Strand R. Rethinking scientific responsibility. *J Med Ethics* 2011;37:299-302.
47. Fugelli P. The Zero-vision: potential side effects of communicating health perfection and zero risk. *Patient Educ Couns* 2006;60:267-71.
48. Gandini S, Sera F, Cattaruzza MS, Pasquini P, Picconi O, Boyle P, Melchi CF. Meta-analysis of risk factors for cutaneous melanoma: II Sun exposure. *Eur J Cancer* 2005;41:45-60.
49. Garbe C, Leiter U. Melanoma epidemiology and trends. *Clinics in Dermatology* 2009;27:3-9.
50. Getz L. Sustainable and responsible preventive medicine. Conceptualizing ethical dilemmas arising from clinical implementation of advancing medical technology. PhD thesis. NTNU: Trondheim 2006. Accessed 4 January 2018. Available at: <http://www.diva-portal.org/ntnu/abstract.xsql?dbid=750>.
51. Gergis A, Campbell EM, Redman S, Sanson-Fisher RW. Screening for melanoma: a community survey of prevalence and predictors. *Med J Australia* 1991;154:338-43.
52. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004;328:39-41.
53. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, Michie S, Moher D, Wager E. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 2014;383:267-76.
54. Goldenberg AJ, Comeau AM, Grosse SD, Tanksley S, Prosser LA, Ojodu J, Botkin JR, Kemper AR, Green NS. Evaluating harms in the assessment of net benefit: a framework for newborn screening condition review. *Matern Child Health J* 2016;20:693-700.
55. Greenhalgh T, Snow R, Ryan S, Rees S, Salisbury H. Six 'biases' against patients and carers in evidence-based medicine. *BMC Med* 2015;13:200.
56. Guirguis-Blake JM, Sun X, Beil TL, Senger CA, Whitlock EP. Primary Care Screening for Abdominal Aortic Aneurysm: A Systematic Evidence Review for the U.S. Preventive Services Task Force [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 Jan. Report No.: 14-05202-EF-1.
57. Gotzsche PC, Hartling OJ, Nielsen M, Brodersen J, Jørgensen KJ. Breast screening: the facts – or maybe not. *BMJ* 2009;338:b86.
58. Gotzsche PC, Jørgensen KJ. The breast screening programme and misinforming the public. *J R Soc Med* 2011;104:361-9.
59. Hakama M, Auvinen A, Day NE, Miller AB. Sensitivity in cancer screening. *J Med Screen* 2007;14:174-7.
60. Hakama M, Pokhrel A, Malila N, Hakulinen T. Sensitivity, effect and overdiagnosis in screening for cancers with detectable pre-invasive phase. *Int J Cancer* 2015;136:928-35.
61. Hansson A, Brodersen J, Reventlow S, Pettersson M. Opening Pandora's box: The experiences of having an asymptomatic aortic aneurysm under surveillance. *Health, Risk & Society* 2012;14,4:341-59.

62. Harach HR, Franssila KO, Wasenius V. Occult papillary carcinoma of the thyroid: a "normal" finding in Finland. A systematic autopsy study. *Cancer* 1985;56:531-8.
63. Haraway D. Situated knowledges: the science question in feminism and the privilege of partial perspective. In: Haraway D, ed. *Simians, cyborgs, and women. The reinvention of nature.* New York: Routledge, 1991:182e201.
64. Harris RP, Sheridan S, Kinsinger L. Time to rethink screening for abdominal aortic aneurysm? *Arch Intern Med* 2012;172:1462-3.
65. Harris RP, Sheridan SL, Lewis CL, Barclay C, Vu MB, Kistler CE, Golin CE, DeFrank JT, Brewer NT. The harms of screening: a proposed taxonomy and application to lung cancer screening. *JAMA Intern Med* 2014;174:281-5.
66. Harris RP, Wilt TJ, Quaseem A, High Value Task Force of the American College of Physicians. A value framework for cancer screening: advice for high-value care from the American College of Physicians. *Ann Intern Med* 2015;162:712-7.
67. Heckman JJ. Skill formation and the economics of investing in disadvantaged children. *Science* 2006;31:1900-2.
68. Heleno B, Thomsen MF, Rodrigues DS, Jørgensen KJ, Brodersen J. Quantification of harms in cancer screening trials: literature review. *BMJ* 2013;347:5334.
69. Hersch J, Barrat A, Jansen J, Irwig L, McGeechan K, Jacklyn G, Thornton H, Dhillon H, Houssami N, McCaffery K. Use of a decision aid including information on overdiagnosis to support informed choice about breast cancer screening: a randomised controlled trial. *Lancet* 2015;385:1642-52.
70. Heywood A, Sanson-Fisher R, Ring I, Mudge P. Risk prevalence and screening for cancer by general practitioners. *Preventive Medicine* 1994;23:152-9.
71. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available at: [www.handbook.cochrane.org](http://www.handbook.cochrane.org)
72. Hoffmann TC, Légaré F, Simmons MB, McNamara K, Trevena LJ, Hudson B, Glasziou PP, Del Mar CB. Shared decision making: what do clinicians need to know and why should they bother? *Med J Aust* 2014;201:35-9.
73. Hoffmann TC, Del Mar C. Patients' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. *JAMA Intern Med* 2015;175:274-86.
74. Hofmann BM. Too much technology. *BMJ* 2015;350:h705.
75. Holford TR, Meza R, Warner KE, Meernik C, Jeon J, Moolgavkar SH, Levy DT. Tobacco control and the reduction in smoking-related premature deaths in the United States, 1964-2012. *JAMA* 2014;311:164-71.
76. Horton R. Offline: What is medicine's 5 sigma? *Lancet* 2015;385:1380.
77. Illich I. Medical Nemesis. *Lancet* 1974;1:918-21.
78. Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, Schulz KF, Tibshirani R. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 2014;383:166-75.
79. Janda M, Elwood M, Ring IT, Firman DW, Lowe JB, Youl PH, Aitken JF. Prevalence of skin screening by general practitioners in regional Queensland. *Med J Aust* 2004;180:10-5.
80. Johansson M, Jørgensen KJ, Marklund B, Hansson A, Brodersen J. Bristfällig rapport om screening för bukaortaaneurysm – gör om!. *Läkartidningen* 2016;113:DU6S.
81. Johansson M, Jørgensen KJ, Marklund B, Hansson A, Brodersen J. Krafttag krävs mot intellektuell bias. *Läkartidningen* 2016;113:EACF.
82. Johansson M, Jørgensen KJ, Marklund B, Hansson A, Brodersen J. Slutreplik: Oroväckande inställning från Socialstyrelsen och SBU. *Läkartidningen* 2016;113:ECRE.
83. Järvinen TLN, Michaëlsson K, Jokihäärä J, Collins GS, Perry TL, Mintzes B, Musini V, Erviti J, Gorricho J, Wright JM, Sievänen H. Overdiagnosis of bone fragility in the quest to prevent hip fracture. *BMJ* 2015;350:h2088.
84. Jørgensen KJ, Klahn A, Gotzsche PC. Are benefits and harms in mammography screening given equal attention in scientific articles? A cross-sectional study. *BMC Med* 2007;5:12.
85. Jørgensen KJ. Melanoma Screening. *Dermatology* 2017;233:410-1.

86. Jørgensen KJ, Johansson M. Drawing conclusions from the VIVA trial. 2018, In Press, The Lancet.
87. Katalinic A, Eisemann N, Waldmann A. Skin cancer screening in Germany. Documenting melanoma incidence and mortality from 2008 to 2013. *Deutsches Arzteblatt International* 2015;112:629-34.
88. Katanoda K. Neuroblastoma mass screening – what can we learn from it? *J Epidemiol* 2016;26:163-5.
89. Keller L, Engelmark M, Holst J, Gottsäter A, Adolfsson J, Hultgren R, Svensjö S, Wanhainen A. Brett stöd för rekommendation om screening för bukaortaaneurysm. *Läkartidningen* 2016;113:D4EH.
90. Keller L, Adolfsson J, Hall C, Larsson LT. Kraven på opartiskhet och vetenskaplighet är höga. *Läkartidningen* 2016;113:ECE4.
91. Kelly MP, Heath I, Howick J, Greenhalgh T. The importance of values in evidence-based medicine. *BMC Med Ethics* 2015;16:69.
92. Kim LG, Scott RAP, Ashton HA, Thompson SG. A sustained mortality benefit from screening for abdominal aortic aneurysm. *Ann Intern Med* 2007;146:699-706.
93. Kirkengen AL, Ekeland TJ, Getz L, Hetlevik I, Schei E, Ulvestad E, Vetlesen AJ. Medicine's perception of reality - a split picture: critical reflections on apparent anomalies within the biomedical theory of science. *J Eval Clin Pract* 2016;22:496-501.
94. Kolthoff SK, Hestbech MS, Jørgensen KJ, Brodersen J. Do invitations for cervical screening provide sufficient information to enable informed choice? A cross-sectional study of invitations for publicly funded cervical screening. *J R Soc Med* 2016;109:274-81.
95. Lakhani NA, Saraiya M, Thompson TD, King SC, Guy GP Jr. Total body skin examination for skin cancer screening among U.S. adults from 2000 to 2010. *Preventive Medicine* 2014;61:75-80.
96. Larsson LT, Engelmark M. Negativa effekter av screening har vägts mot möjliga hälsovinster. *Läkartidningen* 2016;113:DWD3.
97. Le Couteur DG, Doust J, Creasey H, Brayne C. Political drive to screen for pre-dementia: not evidence based and ignores the harms of diagnosis. *BMJ* 2013;347:f5125.
98. Lee CI, Etzioni R. Missteps in current estimates of cancer overdiagnosis. *Acad Radiol* 2017;24:226-9.
99. Lindholt JS, Juul S, Fasting H, Henneberg EW. Preliminary ten year results from a randomised single centre mass screening trial for abdominal aortic aneurysm. *Eur J Vasc Endovasc Surg* 2006;32:608-14.
100. Lindholt JS, Green A, Søgaard R. Screening for abdominal aortic aneurysms has a positive clinical effect. *Ugeskr Læger* 2013;175:2170-4.
101. Lindholt JS, Søgaard R. Population screening and intervention for vascular disease in Danish men (VIVA): a randomised controlled trial. *Lancet* 2017;390:2256-65.
102. Lindholt JS, Søgaard R. Supplementary appendix to: Population screening and intervention for vascular disease in Danish men (VIVA): a randomised controlled trial. *Lancet* 2017. Accessed 4 January 2018. Available at: [http://dx.doi.org/10.1016/s0140-6736\(17\)32250-X](http://dx.doi.org/10.1016/s0140-6736(17)32250-X)
103. Linné A, Smidfelt K, Langenskiöld M, Hultgren R, Nordanstig J, Kragsterman B, Lindström D. Low post-operative mortality after surgery on patients with screening-detected abdominal aortic aneurysms: a Swedvasc registry study. *Eur J Vasc Endovasc Surg* 2014;48:649-56.
104. Lock M, Gordon D. Biomedicine examined. Dordrecht: Kluwer Academic Publishers, 1988.
105. Loewenstein GF, Weber EU, Hsee CK, Welch N. Risk as feelings. *Psychol Bull* 2001;127:267-86.
106. Ma J, Ward EM, Smith R, Jemal A. Annual number of lung cancer deaths potentially avertable by screening in the United States. *Cancer* 2013;119:1381-5.
107. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, Al-Shahi Salman R, Chan AW, Glasziou P. Biomedical research: increasing value, reducing waste. *Lancet* 2014;383:101-4.
108. Marks R. Epidemiology of melanoma. *Clinical & Experimental Dermatology* 2000;25:459-63.

109. Marmot M, Allen J, Bell R, Bloomer E, Goldblatt P, Consortium for the European Review of Social Determinants of Health and the Health divide. WHO European review of social determinants of health and the health divide. *Lancet* 2012;380:1011-29.
110. Martin SA, Boucher M, Wright JM, Saini V. Mild hypertension in people at low risk. *BMJ* 2014;349:g5432.
111. Maund E, Tendal B, Hróbjartsson A, Jørgensen KJ, Lundh A, Schroll J, Gøtzsche PC. Benefits and harms in clinical trials of duloxetine for treatment of major depressive disorder: comparison of clinical study reports, trial registries, and publications. *BMJ* 2014;348:g3510.
112. McCaffery KJ, Barrat AL. Assessing psychosocial/quality of life outcomes in screening: how do we do it better? *J Epidemiol Community Health* 2004;58:968-70.
113. McCaul KA, Lawrence-Brown M, Dickinson JA, Norman PE. Long-term Outcomes of the Western Australian Trial of Screening for Abdominal Aortic Aneurysms: Secondary Analysis of a Randomized Clinical Trial. *JAMA Intern Med* 2016;176:1761-7.
114. Meixel A, Yanchar E, Fugh-Berman A. Hypoactive sexual desire disorder: inventing a disease to sell low libido. *J Med Ethics* 2015;41:859-62.
115. Miller MR, Levy ML. Chronic obstructive pulmonary disease: missed diagnosis versus misdiagnosis. *BMJ* 2015;351:h3021.
116. Moyer VA, U.S. Preventive Services Task Force. Screening for cervical cancer: U.S. Preventive Services recommendation statement. *Ann Intern Med* 2012;156:880-91.
117. Moyer VA, U.S. Preventive Services Task Force. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2014;160:330-8.
118. Moynihan R, Glasscock R, Doust J. Chronic kidney disease controversy: how expanding definitions are unnecessarily labelling many people as diseased. *BMJ* 2013;347:f4298
119. Munn Z, Porritt K, Lockwood C, Aromataris E, Pearson A. Establishing confidence in the output of qualitative research synthesis: the ConQual approach. *BMC Med Res Methodol* 2014;14:108.
120. Nelson WL, HanPK, Fagerlin A, Stefanek M, Ubel PA. Rethinking the objectives of decision aids: a call for conceptual clarity. *Med Decis making* 2007;27:609-18.
121. Norgaard C, Glud M, Gniadecki R. Are all melanomas dangerous? *Acta Dermato-Venerologica* 2011;91:499-503.
122. OECD (2015), Fiscal Sustainability of Health Systems: Bridging Health and Finance Perspectives, OECD Publishing, Paris. Accessed 4 January 2018. Available at: <http://dx.doi.org/10.1787/9789264233386-en>
123. Office for National Statistics. Opinions and lifestyle survey, Smoking habits amongst adults 2012. Accessed 4 January 2018. Available at: [www.ons.gov.uk/ons/dcp171776\\_328041.pdf](http://www.ons.gov.uk/ons/dcp171776_328041.pdf)
124. Ohlsson H, Gottsäter A, Resch T, Sonesson B, Malina M Kjellin P, Wetterling T, Holst J. On the complexity of screening detected abdominal aortic aneurysms: a retrospective observational multicentre cohort study. *Int Angiol* 2017;36:261-7.
125. Pettersson M, Mattson E, Bergbom I. Prospective follow-up of sexual function after elective repair of abdominal aortic aneurysms using open and endovascular techniques. *J Vasc Surg* 2009;50:492-9.
126. Pétursson H. The validity and relevance of international cardiovascular disease prevention guidelines for general practice. PhD thesis. NTNU: Trondheim 2012. Accessed 4 January 2018. Available at: <https://brage.bibsys.no/xmlui/handle/11250/264486>
127. Ploug T, Holm S, Brodersen J. To nudge or not to nudge: cancer screening programmes and the limits of libertarian paternalism. *J Epidemiol Community Health* 2012;66:1193-6.
128. Prasad V, Lenzer J, Newman DH. Why cancer screening has never been shown to "save lives"—and what we can do about it. *BMJ* 2016;352:h6080.
129. Raffle A, Gray M. Screening - evidence and practice. New York: Oxford University Press, 2007.
130. Rasmussen JF, Siersma V, Pedersen JH, Heleno B, Saghir Z, Brodersen J. Healthcare costs in the Danish randomised controlled lung cancer CT-screening trial: a registry study. *Lung Cancer* 2014;83:347-55.
131. Reventlow SD, Hvas L, Malterud K. Making the invisible body visible. Bone scans, osteoporosis and women's bodily experiences. *Soc Sci Med* 2006;62:2720-31.

132. Sackett DL. The arrogance of preventive medicine. *CMAJ* 2002;167:363-4.
133. Sakr WA, Grignon DJ, Haas GP, Heilbrun LK, Pontes JE, Crissman JD. Age and racial distribution of prostatic intraepithelial neoplasia. *Eur Urol* 1996;30:138-44.
134. Saraiya M, Hall HI, Thompson T, Hartman A, Glanz K, Rimer B, Rose D. Skin cancer screening among U.S. adults from 1992, 1998, and 2000 National Health Interview Surveys. *Preventive Medicine* 2004;39:308-14.
135. Sawada T, Todo S, Fujita K, Iino S, Imashuku S, Kusunoki T. Mass screening of neuroblastoma in infancy. *Am J Dis Child* 1982;136:710-2.
136. Schanzer A, Greenberg RK, Hevelone N, Robinson WP, Eslami MH, Goldberg RJ, Messina L. Predictors of abdominal aortic aneurysm sac enlargement after endovascular repair. *Circulation* 2011;123:2848-55.
137. Schilling FH, Spix C, Berthold F, Erttmann R, Fehse N, Hero B, Klein G, Sander J, Schwarz K, Treuner J, Zorn U, Michaelis J. Neuroblastoma screening at one year of age. *N Engl J Med* 2002;346:1047-53.
138. Schroll JB, Penninga EI, Gotzsche PC. Assessment of Adverse Events in Protocols, Clinical Study Reports, and Published Papers of Trials of Orlistat: A Document Analysis. *PLoS Med* 2016;13:e1002101.
139. Schwartz LM, Woloshin S, Welch HG. Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. *N Engl J Med* 1999;34:279-83.
140. Schwartz LM, Woloshin S, Fowler FJ Jr, Welch HG. Enthusiasm for cancer screening in the United States. *JAMA* 2004;29:71-8.
141. Shaw RL, Larkin M, Flowers P. Expanding the evidence within evidence-based healthcare: thinking about the context, acceptability and feasibility of interventions. *Evid Based Med* 2014;19:201-3.
142. Shickle D, Chadwick R. The ethics of screening: is 'screeninggitis' an incurable disease? *J Med Ethics* 1994;20:12-8.
143. Slovic P, Finucane M, Peters E, MacGregor DG. The Affect Heuristic. In: Gilovich T, Griffin D, Kahneman D, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press, 2002:397-420.
144. Stamatiou K, Alevizos A, Agapitos E, Sofras F. Incidence of impalpable carcinoma of the prostate and of non-malignant and precarcinomatous lesions in the Greek male population: an autopsy study. *Prostate* 2006;66:1319-28.
145. Stather PW, Sidloff D, Dattani N, Choke E, Bown MJ, Savers RD. Systematic review and meta-analysis of the early and late outcomes of open and endovascular repair of abdominal aortic aneurysm. *Br J Surg* 2013;100:863-72.
146. Svensjö S, Björck M, Gurtelschmid M, Djavani GK, Hellberg A, Wanhainen A. Low prevalence of abdominal aortic aneurysm among 65-year-old Swedish men indicates a change in the epidemiology of the disease. *Circulation* 2011;124:1118-23.
147. Svensjö S, Björck M, Wanhainen A. Current prevalence of abdominal aortic aneurysm in 70-year-old women. *Br J Surg* 2013;100:367-72.
148. Sängren H, Reventlow S, Hetlevik I. Role of biographical experience and bodily sensations in patients' adaptation to hypertension. *Patient Educ Couns* 2009;74:236-43.
149. Thompson SG, Ashton HA, Gao L, Scott RA. Screening men for abdominal aortic aneurysm: 10 year mortality and cost effectiveness results from the randomised multicentre aneurysm screening study. *BMJ* 2009;338:b2307.
150. Thompson AR, Cooper JA, Ashton HA, Hafez H. Growth rates of small abdominal aortic aneurysms correlate with clinical events. *Br J Surg* 2010;97:37-44.
151. Thompson SG, Ashton HA, Gao L, Buxton MJ, Scott RAP. Final follow-up of the Multicentre Aneurysm Screening Study (MASS) randomized trial of abdominal aortic aneurysm screening. *Br J Surg* 2012;99:1649-56.
152. Tómasdóttir MO. Multimorbidity in the Norwegian HUNT population. PhD thesis. NTNU: Trondheim 2017. Accessed 4 January 2018. Available at: <https://opnvisindi.is/bitstream/handle/20.500.11815/473/Thesis-finalMOT.pdf?sequence=1&isAllowed=y>

153. UK National Screening Committee. Criteria for appraising the viability, effectiveness and appropriateness of a screening programme. Accessed 2 June 2015. Available at: [www.screening.nhs.uk/criteria](http://www.screening.nhs.uk/criteria)
154. Ulug P, Powell JT, Sweeting MJ, Bown MJ, Thompson SG. Meta-analysis of the current prevalence of screen-detected abdominal aortic aneurysm in women. *Br J Surg* 2016;103:1097-104.
155. Ulug P, Sweeting MJ, von Allmen RS, Thompson SG, Powell JT, on behalf of the SWAN collaborators. Morphological suitability for endovascular repair, non-intervention rates, and operative mortality in women and men assessed for intact abdominal aortic aneurysm repair: systematic reviews with meta-analysis. *Lancet* 2017;389:2482-91.
156. Vardulaki KA, Prevost TC, Walker NM, Day NE, Wilmink ABM, Quick CRG, et al. Growth rates and risk of rupture of abdominal aortic aneurysms. *Br J Surg* 1998;85:1674-80.
157. Vogt H. Systems Medicine as a Theoretical Framework for Primary Care Medicine. PhD thesis. NTNU: Trondheim 2017. Accessed 4 January 2018. Available at: [https://brage.bibsys.no/xmlui/bitstream/handle/11250/2441891/Henrik%20Vogt\\_PhD.pdf?sequence=1](https://brage.bibsys.no/xmlui/bitstream/handle/11250/2441891/Henrik%20Vogt_PhD.pdf?sequence=1)
158. Vänni V, Turtiainen J, Hakala T, Salenius J, Suominen V, Oksala N, Hernesniemi J. Vascular comorbidities and demographics of patients with ruptured abdominal aortic aneurysms. *Surgery* 2016;159:1191-8.
159. Waldmann A, Nolte S, Weinstock MA, Breitbart EW, Eisemann N, Geller AC, Greinert R, Volkmer B, Katalinic A. Skin cancer screening participation and impact on melanoma incidence in Germany – an observational study on incidence trends in regions with and without population-based screening. *British J Cancer* 2012;106:970-4.
160. Wanhainen A, Björck M. The Swedish experience of screening for abdominal aortic aneurysm. *J Vasc Surg* 2011;53:1164-5.
161. Wegwarth O, Schwartz LM, Woloshin S, Gaissmaier W, Gigerenzer G. Do physicians understand cancer screening statistics? A national survey of primary care physicians in the United States. *Ann Intern Med* 2012;156:340-9.
162. Weiss NS, Rossing MA. Healthy screened bias in epidemiologic studies of cancer incidence. *Epidemiology* 1996;7:319-22.
163. Welch HG, Woloshin S, Schwartz LM. Skin biopsy rates and incidence of melanoma – population based ecological study. *BMJ* 2005;331:481.
164. Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst* 2010;102:605-13.
165. Welch HG, Schwartz L, Woloshin S. Overdiagnosed—making people sick in the pursuit of health. Boston: Beacon Press, 2011.
166. Welch HG, Frankel BA. Likelihood that a woman with screen-detected breast cancer has had her "life saved" by that screening. *Arch Intern Med* 2011;171:2043-6.
167. Wernli KJ, Henrikson NB, Morrison CC, Nguyen M, Pocobelli G, Whitlock EP. Screening for skin cancer in adults: An updated systematic evidence review for the U.S. preventive services task force. Evidence Synthesis No. 137. AHRQ Publication No. 14-05210-EF-1. Rockville, MD: Agency for Healthcare Research and Quality 2016.
168. Wiener RS, Schwartz LM, Woloshin S. When a test is too good: how CT pulmonary angiograms find pulmonary emboli that do not need to be found. *BMJ* 2013;347:f3368.
169. Woloshin S, Schwartz LM, Black WC, Kramer BS. Cancer screening campaigns—getting past uninformative persuasion. *N Engl J Med* 2012;367:1677-9.
170. Woods WG, Gao RN, Shuster JJ, Robison LL, Bernstein M, Weitzman S, Bunin G, Levy I, Brossard J, Dougherty G, Tuchman M, Lemieux B. Screening of infants and mortality due to neuroblastoma. *N Engl J Med* 2002;346:1041-6.
171. Wrede A, Wiberg F, Acosta S. Increasing the elective endovascular to open repair ratio of popliteal artery aneurysm. *Vasc Endovasc Surg* 2018;52:115-23.
172. Yudkin JS, Montori VM. The epidemic of pre-diabetes: the medicine and the politics. *BMJ* 2014;349:g4485.
173. Zahl PH, Mæhlen J, Welch HG. The natural history of invasive breast cancers detected by screening mammography. *Arch Intern Med* 2008;168:2311-6

174. Zahl PH, Jørgensen KJ, Gotzsche PC. Overestimated lead times in cancer screening has led to substantial underestimation of overdiagnosis. *Br J Cancer* 2013;109:2014-9.
175. Ziganshin BA, Elefteriades JA. Guilt by association: a paradigm for detection of silent aortic disease. *Ann Cardiothorac Surg* 2016;5:174-87.