Reducing Harms and Increasing Benefits of Screening for Cervical Cancer and Colorectal Cancer

A Model-based Approach

Steffie Naber

Reducing Harms and Increasing Benefits of Screening for Cervical Cancer and Colorectal Cancer – A Model-based Approach

Steffie Katinka Naber

Dit proefschrift is mede tot stand gekomen dankzij de financiële steun van de afdeling Maatschappelijke Gezondheidszorg van het Erasmus MC.

Layout and printing: Optima Grafische Communicatie, Rotterdam, the Netherlands
Cover design: Robert Dumay Graphic Design

ISBN: 978-94-92683-14-4

Reducing Harms and Increasing Benefits of Screening for Cervical Cancer and Colorectal Cancer – A Model-based Approach

Het verkleinen van schadelijke effecten en vergroten van gunstige effecten van screening op baarmoederhalskanker en darmkanker – een modelmatige aanpak

Proefschrift

Ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
woensdag 19 april 2017 om 11.30 uur

door

Steffie Katinka Naber
geboren te Zoetermeer

Erasmus University Rotterdam

Erasmus

PROMOTIECOMMISSIE

Promotor:	Prof.dr. H.J. de Koning
Overige leden:	Prof.dr. F.J. van Kemenade Prof.dr. C.M. van Duijn Prof.dr. C.J.L.M. Meijer
Copromotoren:	Dr. I. Lansdorp-Vogelaar Dr. I.M.C.M. de Kok

CONTENTS

Chapter 1

General introduction

WHAT IS SCREENING?

Screening refers to the systematic testing in asymptomatic individuals to identify disease or risk factors for disease. This enables the possibility to act earlier, e.g. by starting treatment of the disease earlier. Screening is especially valuable for diseases like cancer, where prognosis is better when treated in earlier stages.

In the Netherlands, organized screening programs exist for early detection of cervical cancer, breast cancer, and colorectal cancer. Target populations are invited systematically, and follow-up of those with a positive screening test is ensured. The quality of each step in the screening process is measured, reported, and evaluated. As opposed to such organized screening, opportunistic screening is delivered on an ad hoc basis, where uptake depends on requests from individuals or recommendations from health care providers. Due to lack of organization, opportunistic screening is at greater risk for overuse and underuse of screening resources.[1]

Screening for disease aims to identify individuals that will be diagnosed with that disease within the (near) future, and for whom earlier detection and treatment is expected to reduce morbidity and mortality from that disease. However, even if early detection and treatment is expected to decrease the burden of a specific disease, it is not necessarily suitable for mass screening. Although some may have an idealized view of screening, the effects are never solely positive. Any type of screening program comes with burden (e.g. primary testing and waiting for the result), harms (e.g. complications of treatment of pre-invasive lesions), and significant costs. Therefore, there are several criteria that should be met prior to implementation of mass screening.

CRITERIA FOR SCREENING

In 1968, the World Health Organization (WHO) commissioned the report "Principles of early disease detection", in which Wilson and Jungner proposed a first set of requirements for population-based screening (**Box 1**).[2] Although the authors merely hoped that their publication would stimulate discussion, their criteria are still regarded as gold standard for the evaluation of screening programs. Forty years later, after many screening programs had been implemented, the WHO criteria were updated.[3]

There is general consensus that prior to implementation of screening, all of the (revised) WHO criteria should be met. The section of Early Detection at the Department of Public Health at the Erasmus MC suggests the use of a two-step approach, starting with a more condensed set of criteria:

1. There should be scientific evidence of screening program effectiveness;
2. The overall benefits of screening should outweigh the harm;
3. There should be scientific evidence of screening program cost-effectiveness.

Box 1. WHO criteria for screening.

Criteria proposed by Wilson and Jungner (1968)

- The condition sought should be an important health problem.
- There should be an accepted treatment for patients with recognized disease.
- Facilities for diagnosis and treatment should be available.
- There should be a recognizable latent or early symptomatic stage.
- There should be a suitable test or examination.
- The test should be acceptable to the population.
- The natural history of the condition, including development from latent to declared disease, should be adequately understood.
- There should be an agreed policy on whom to treat as patients.
- The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.
- Case-finding should be a continuing process and not a "once and for all" project.

Revised criteria proposed by Andermann et al. (2008)

- The screening program should respond to a recognized need.
- The objectives of screening should be defined at the outset.
- There should be a defined target population.
- There should be scientific evidence of screening program effectiveness.
- The program should integrate education, testing, clinical services and program management.
- There should be quality assurance, with mechanisms to minimize potential risks of screening.
- The program should ensure informed choice, confidentiality and respect for autonomy.
- The program should promote equity and access to screening for the entire target population.
- Program evaluation should be planned from the outset.
- The overall benefits of screening should outweigh the harm.

In case screening does not satisfy one or more of these criteria, screening should not be considered and further evaluation of the WHO criteria is not necessary. Alternatively, if all three criteria are met, a country or organization can proceed with evaluating the full set of WHO criteria to determine whether screening is feasible given its specific needs and resources. The next sections describe the aforementioned three criteria for screening in more detail.

1. Proof of effectiveness

Scientific evidence should show that screening is effective in reducing the disease burden. Ideally, the effectiveness of screening is shown in a randomized controlled trial (RCT) with two (or more) study arms: one with screening (intervention arm), and one without screening (control arm). Primary outcome of such a trial should be disease-specific incidence or mortality. If, in the intervention arm, significantly fewer individuals are diagnosed with the disease or die from the disease, then screening reduces the morbidity or mortality from that disease, and it can be considered effective.

Compared to other study designs, the major benefit of an RCT is the fact that individuals are randomly allocated to either the intervention or the control arm. This randomization ensures that the distribution of baseline risk is similar in both arms. Drawbacks of RCTs include that they are expensive and time-consuming. Alternatively, non-randomized studies can be used to estimate the effectiveness of screening. In these studies, however, differences in

health outcomes between attendees and non-attendees of screening may be the result of differences in disease risk at baseline. It is difficult to measure the effectiveness of screening if such a selection bias is present.

The disease-specific incidence or mortality as a primary outcome of an RCT is very important. Another outcome that is often reported is the difference in survival between screen-detected and clinically diagnosed cases. This comparison, however, is prone to two other forms of bias, i.e. lead-time and length-time bias.

Lead time refers to the time frame from the moment of cancer detection by screening until the moment that the cancer would have been clinically diagnosed in the absence of screening. Lead-time bias occurs when this time frame is regarded as prolonged survival for screen-detected cancers, whereas in fact it only concerns a shift from healthy life years to life years lived with clinical disease (**Figure 1**).

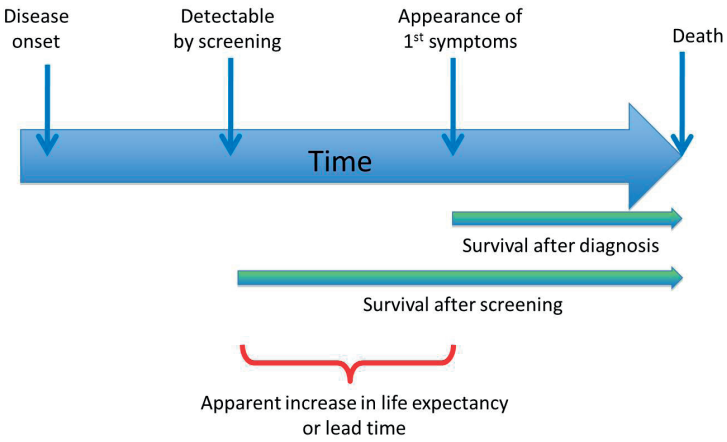


Figure 1. Visualization of lead-time bias.[4]

Length-time bias refers to the phenomenon that cancers with a long preclinical duration are more likely to be detected with screening than those with a short preclinical duration. For slow progressive disease (i.e. cancers with a long preclinical duration), the time from the onset of disease until the death from disease is longer than for more rapidly progressive disease (i.e. cancers with a short preclinical duration). Even if early detection of cancer does not improve survival probabilities, screen-detected cancers have a longer survival time because screening is more likely to detect slow-growing disease (**Figure 2**).

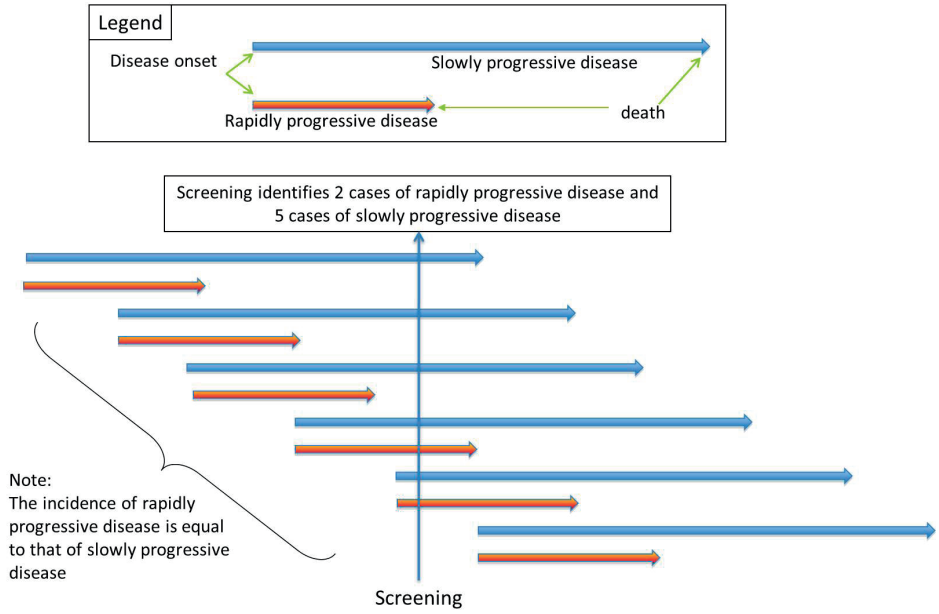


Figure 2. Visualization of length-time bias.[4]

Due to lead-time and length-time bias, screen-detected cancers (seem to) have a more favorable survival, and screening may seem very effective. To ensure that the effect of screening is not being overestimated, the survival found in RCTs should therefore be corrected for these two types of bias.

2. Benefits should outweigh the harm

Once the effectiveness of a program has been proven, one should determine whether the overall benefits outweigh the harm. Often, many individuals have to be screened to prevent one individual from being diagnosed with (or dying from) the disease. In general, it is difficult to judge whether this benefit of screening for one person is worth the burden of screening for all the others. It requires an estimate of the impact on quality of life of the screening itself, and of the different screening outcomes, both of which may differ substantially from one person to the other.

Screening

Screening itself can be experienced as burdensome. Receiving an invitation, and being aware of the fact that you might have (a pre-stadium of) the disease, for some individuals already has a negative impact on their quality of life. They are forced to think about the possibility of them having the disease, and about whether they should participate in screening. For some types of screening, the test itself can be experienced as burdensome or even painful, like a mammogram in breast cancer screening. After testing, participants

have to wait for the result, and during this period participants might suffer from feelings of anxiety and stress. Individuals with a positive test result are either treated directly (as with colonoscopy screening) or, in most cases, are referred for follow-up testing. Higher levels of anxiety and stress will be experienced towards this second, diagnostic test.

Screening outcomes

Screening tests can be either positive or negative, and this may either be correctly identifying the absence or presence of disease, or it may be misclassifying someone as healthy or ill. **Table 1** provides an overview of the benefits and harms related to those different screening outcomes.

Table 1. Harms and benefits of screening outcomes.

		Test result	
		Positive	Negative
(Precursor of) disease	Present	**True positive** Benefits: Early detection of progressive disease, thereby enabling early treatment and reducing morbidity or mortality from disease Harms: Detection and treatment of non-progressive disease, i.e. overdiagnosis and overtreatment	**False negative** Benefits: -- Harms: False reassurance, potentially leading to a later clinical diagnosis
	Absent	**False positive** Benefits: -- Harms: Unnecessary follow-up testing, and (additional) anxiety and stress	**True negative** Benefits: Justified reassurance Harms: --

True positive test result

Clearly, the major benefit of screening for disease is the early detection of progressive disease, thereby enabling early treatment. This requires a true positive test result at screening. However, the detected preclinical disease might have never developed into clinical disease in the absence of screening. An individual only benefits from early detection if the earlier initiation of treatment prevents morbidity or mortality from the disease. By detecting preclinical disease that would never have progressed to clinical cancer, individuals are being overdiagnosed and potentially also overtreated. Although the severity of overtreatment hugely depends on (the stage of) the disease, it will surely have a negative impact on someone's quality of life.

False positive test result

A screening test might also produce a positive test result in individuals without (a precursor of) the disease. In such case, individuals will face unnecessary worry, stress and follow-up testing, which potentially reduces their quality of life. Although the negative test result at follow-up will show that any increased concerns were unnecessary, individuals might not be entirely reassured.

True negative test result
A negative test result at baseline is more likely to reassure individuals. Those who test negative and do not have (a precursor of) the disease, may benefit from justified reassurance.

False negative test result
With any type of screening, some prevalent disease will be missed. Individuals with a negative test result, but with (a precursor of) the disease, are falsely reassured by the screening. This may lead to a delayed clinical diagnosis, as symptoms may be recognized in a later stage.

3. Proof of cost-effectiveness

If studies have unequivocally demonstrated that the benefits of screening outweigh the harms, then the costs of the screening program should be evaluated. Screening programs are expensive, and health care budgets can only be spent once. If the same budget can save more life years in preventing or treating another disease, then, from a cost-effectiveness perspective, budgets should be directed to that disease first.

The cost-effectiveness of interventions can be evaluated in cost-effectiveness analyses (CEAs), which can be performed using mathematical models. In screening CEAs, strategies are compared in terms of their costs and (quality-adjusted) life years. Strategies are deemed cost-effective if they are considered to be good value for money, given alternative options. More formally, the additional cost per additional (quality-adjusted) life year gained of one strategy versus another is computed, and compared with the amount that society is willing to pay for one (quality-adjusted) life year. Although it seems straightforward that in order to make such a comparison, all relevant alternatives should be included in an analysis, it is not always straightforward what exactly comprises a relevant strategy. This is one of the issues we aim to address in **Chapter 2** of this thesis, where we have reviewed the choice of strategies included in CEAs of HPV-based cervical cancer screening.

CERVICAL CANCER AND COLORECTAL CANCER

This thesis includes studies on both cervical and colorectal cancer screening. The natural history of these diseases is similar in the sense that a long preclinical phase precedes cancer development. During this preclinical phase, benign lesions can be detected by screening. Although any treatment is associated with risk of complications, most precursor lesions of cervical and colorectal cancer can be removed relatively easy, thereby preventing cancer development. Alternatively, screening can detect cancer while it is still asymptomatic. For both cervical and colorectal cancer, prognosis steeply increases when detected in an earlier stage. For cases diagnosed in 2005-2011 in the United States, five-year survival of distant cancer was only 17% for cervical and 12% for colorectal cancer, whereas five-year survival of localized cancer was 92% for cervical and 90% for colorectal cancer.[4] On the one hand, this

makes both diseases extremely suitable for screening. On the other hand, benign lesions in the cervix and colon are common, and screening can result in overdiagnosis and overtreatment of such lesions. In order to minimize this potential harm of screening, over-screening (i.e. screening more often than recommended) should be kept to a minimum. Per-treatment complication risks might be low, but together with psychological distress, the harms of screening can outweigh the benefits for individuals who are intensively screened.

The epidemiology, natural history, risk factors and current status of prevention of cervical and colorectal cancer are described in more detail below.

Cervical cancer

Cervical cancer is the fourth most common cancer in women worldwide, with an estimated 528,000 new cases and an estimated 266,000 deaths in 2012.[5] The burden of cervical cancer concentrates in the developing world, where approximately 84% of all new cases are found.[5]

In countries with well-developed screening programs, cervical cancer incidence and mortality have reduced substantially over the past decades. This decline is mainly due to the early detection and treatment of (precursors of) squamous cell carcinomas, which comprise ~90% of incident cases worldwide.[6] The effectiveness of cytological screening in reducing the incidence and mortality of adenocarcinoma (i.e., the other ~10% of incident cases) is less evident.

Natural history

Infection with the human papillomavirus (HPV) is a necessary, but not a sufficient cause of cervical cancer.[7] HPV infections are common sexual transmitted infections (STIs), that are transmitted via skin-to-skin contact. Although most HPV infections clear naturally, some are persistent and precursor lesions may develop. These lesions, so-called cervical intraepithelial neoplasia (CIN), may transition from grade 1 to grade 2, and from grade 2 to grade 3. Data on the progression of CIN to cancer are limited, because once detected, it is considered unethical to leave the lesions untreated. To date, there has been one study in which CIN 3 was left untreated though. The clinician involved in this study did not believe CIN to be a precursor of cancer. The results have shown that if left untreated for 30 years, 31-50% of CIN 3 lesions develop into clinical cancer.[8] The entire process from acquiring an HPV infection to being diagnosed with cervical cancer is depicted in **Figure 3**, and takes on average approximately 15-20 years.

Although nearly half of CIN 3 lesions are true precursors of cervical cancer, CIN 1 and CIN 2 often regress naturally. Especially for CIN 1, progression to cervical cancer is rare. Several guidelines therefore state that women with CIN 1 should not be treated (directly), but should be offered follow-up testing instead.[9,10] If the lesion has not regressed after 1-3 years, then treatment can still be offered.

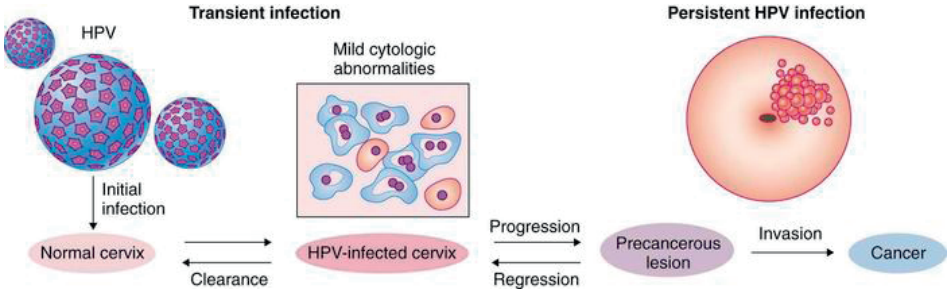


Figure 3. Natural history of cervical cancer.[10]

Risk factors of cervical cancer

Risk factors of cervical cancer can be subdivided into those that increase the probability of acquiring HPV, and those that increase the probability of progression to CIN and cervical cancer.

Probability of acquiring HPV

Just as for all STIs, sexual behavior influences the probability of acquiring an HPV infection. Indeed, age at first sexual intercourse[11] and number of new and recent sexual partners[12] were found to be associated with the risk of acquiring the virus. HPV is presumably transmitted by skin-to-skin contact,[12] and is found in male genital areas not covered by a condom.[13] Although inconsistent data have been published on the protective effect of condoms,[14] the consistent use of condoms appears to be associated with lower rates of acquiring HPV.[15]

Probability of progression to (pre-)invasive disease

In addition to a lower probability of acquiring HPV, women who consistently use condoms have a higher probability of clearing an HPV infection and regressing any present CIN.[15] Although progression to high-grade CIN was found to be positively correlated with cigarette smoking[16,17], long-term oral contraceptive use[17,18], and multiparity[17], there is insufficient evidence as to whether these risk factors are also associated with an increased risk of cancer development.[16,17] To date, there have been no indications that, apart from behavioral differences, ethnic background is associated with cervical cancer risk.[19] Women infected with human immunodeficiency virus (HIV) are at an increased risk of developing both CIN[20,21] and invasive cervical cancer[22]. In addition, compared to women without HIV, the prognosis of HIV-infected women is worse after invasive cervical cancer has been diagnosed.[23] Women carrying the BRCA1 mutation are also at an increased risk of developing cervical cancer.[24]

Current status of HPV vaccination

Since the discovery of HPV being a necessary cause of cervical cancer, HPV vaccines have been developed. Over the past decade, vaccination with either a bivalent or quadrivalent vaccine has been implemented in many developed countries.[25] Both vaccines target the two highly oncogenic HPV-types 16 and 18, together causing roughly 80% of all cervical

cancer cases.[26] Vaccination has shown to be highly effective in reducing the prevalence of HPV-16 and HPV-18.[27,28] Nevertheless, coverage is still relatively low, with only ~40% of all targeted girls being vaccinated worldwide.[25] Therefore, and because other oncogenic HPV-types may still cause cancer, birth cohorts that have been offered routine vaccination should still be invited for cervical cancer screening.

Current status of cervical cancer screening

In most developed countries, cytology-based screening has been in place for decades. With cytology, a smear is taken from the cervix, and is sent to the laboratory, where it is placed under a microscope and screened for abnormalities by a cytologist. Using a standardized scale, the degree of abnormalities is rated. In most settings, women with low-grade abnormalities are invited for triage testing, and women with high-grade abnormalities are referred for colposcopy.

Over the past 15 years, conventional cytology has been replaced by liquid-based cytology (LBC) in several countries, such as the UK, the US, Australia, and the Netherlands. Both conventional cytology and LBC are performed on a cervical smear, which is collected by scraping off cells from the transformation zone of the cervix. The difference between the techniques is that with conventional cytology, the cells are directly smeared on a slide, whereas with LBC, the cells are rinsed in a vial with preservation solution and this vial is then transferred to the laboratory, where a uniform layer of cells is put on a slide. In most cases, the uniform layer does not require the inclusion of all collected cells, and the cells that are left in the vial can be used for HPV testing. This is a major advantage of LBC compared to conventional cytology. The use of LBC was also found to result in fewer smears of unsatisfactory quality.[29] Since the development of a wide range of HPV tests, they were gradually incorporated as a triage test in cytology-based screening programs. In 2017, the Netherlands and Australia will be the first countries to switch from a screening program with primary LBC testing, to one with primary HPV testing.[30,31] In both countries, unscreened and underscreened women will be offered a self-sampling kit, which can be used to self-collect vaginal material at home.[30,32] Just as with the regular HPV test, the collected material can then be sent to the laboratory, where it is tested on the presence of HPV.

Colorectal cancer

Globally, colorectal cancer is the third most common cancer in men, and the second in women, with an estimated 1,360,000 cases and nearly 700,000 deaths in 2012.[5] Colorectal cancer risk increases with age, and is higher in men than in women.[33] In contrast with cervical cancer, colorectal cancer is much more a disease of the developed world, where more than half of all cases are found.[5] The Western lifestyle, including excessive alcohol and red meat consumption, smoking and obesity, significantly increases the probability of developing colorectal cancer.[34]

Natural history

Colorectal cancer is believed to develop from non-malignant polyps in the colon. Until recently, all cancers were believed to develop through the type of non-malignant polyps called adenomas. These adenomas may grow in size, after which they might become malignant. This development is called the adenoma-carcinoma sequence (**Figure 4**). Whereas adenomas are thought to be present in 20-53% of the US population over age 50, the lifetime risk of developing an adenocarcinoma is only 5%.[35] An early study suggested that it takes on average 10-15 years for an adenoma to progress to cancer.[36]

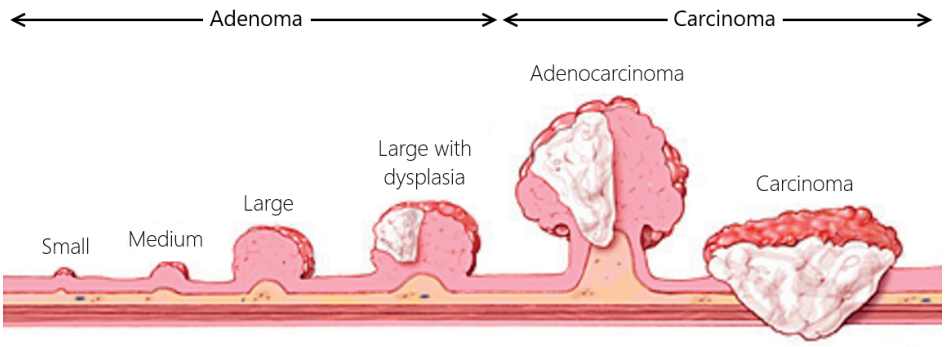


Figure 4. Traditional adenoma-carcinoma sequence.

More recently, a new pathway to colorectal cancer has been identified, called the serrated polyp pathway. This pathway also involves the development of non-malignant polyps, but other than adenomas they are non-neoplastic, but rather sessile serrated polyps. The magnitude of impact of this pathway is subject of debate and currently the most quoted number is that approximately 15% of cancers develop through this pathway.[35] Data on the natural history of this pathway are scant.

Risk factors of colorectal cancer

It has been estimated that more than half of colorectal cancer cases are caused by lifestyle and environmental factors.[37] The largest increase in risk is caused by alcohol consumption, with a 60% higher risk for heavy drinkers compared to non- or light drinkers.[34] Smoking, diabetes, obesity and high meat intakes are independently associated with an increased colorectal cancer risk of 20%.[34] Physical activity and the use of aspirin provide a protective effect.[34,40] The increased risk of bleeding complications in individuals on high-dose aspirin may limit its usefulness in terms of primary prevention of colorectal cancer though.[40]

While most cases of colorectal cancer are sporadic, some can be linked to a specific inherited cancer syndrome, or to other less pronounced genetic factors. Approximately 5% of colorectal cancer diagnoses are found in individuals with a cancer syndrome, such as Lynch Syndrome or Familial Adenomatous Polyposis.[38] As people with a positive family history, but without an identified cancer syndrome, are also at an increased risk for developing the

disease, there must be other inherited factors that increase colorectal cancer risk. Currently a considerable amount of research is devoted to identification of single nucleotide polymorphisms (SNPs). These are mutations that are much more prevalent in the population than the genetic factors that have been linked to cancer syndromes. Although individually, their effect is much less pronounced, people with multiple SNPs can have a substantially increased risk.

Current status of colorectal cancer screening

The status of colorectal cancer screening varies widely across countries.[39] Whereas in the United States, opportunistic screening was already introduced in the 1980s, most European countries have only recently introduced organized screening.[39] This has been triggered by multiple RCTs showing the effectiveness of colorectal cancer screening.[40-50]

Colorectal cancer screening can be performed with a wide range of screening tests, which can be subdivided into stool-based, imaging and endoscopic tests. Stool-based tests include the guaiac fecal occult blood test (gFOBT) and fecal immunochemical test (FIT), which are both widely used in organized screening programs.[39] Stool-based tests require the participant to collect stool and send it to the laboratory, where it is tested for the presence of blood by targeting heme (in case of gFOBT) or human globin (in case of FIT). In contrast with gFOBT, FIT is a quantitative test, meaning that the positivity cut-off level can be adjusted.[39] Because of its limited sensitivity, stool-based testing is often recommended at an annual or biennial basis.[39] One of the advantages of the more invasive imaging and endoscopic tests is that they are more sensitive, and therefore require less frequent testing. The drawback of imaging and endoscopic tests, however, is that they require a bowel preparation to clean out the colon. With imaging techniques, such as computed tomographic colonography (CTC), individuals are also exposed to low dose radiation, which is used to obtain an interior view of the colon. This technique is costlier and more invasive than stool-based testing and its use is much more limited. Most countries with opportunistic screening rely on the use of endoscopy, in particular colonoscopy.[39] With endoscopy, a flexible tube with a fiber-optic camera is inserted into the anus, and pushed through the colon to visualize any adenomas or cancers. A major benefit of this technique is that, in most cases, detected adenomas can directly be removed. Whereas with colonoscopy the colon is fully visualized in ~95% of cases, sigmoidoscopy aims at visualization of the left colon only. Therefore, sigmoidoscopy is less sensitive than colonoscopy, but it is also less burdensome and it involves a lower risk of complications. As colonoscopy is regarded as the gold standard in terms of sensitivity, screening with tests other than colonoscopy require a diagnostic colonoscopy after a positive test result.

Despite the wide range of available screening tests in the US, adherence with colorectal cancer screening is relatively low. Whereas 81% of 21-65 year-old women had a cervical smear taken in the past 3 years, only 59% of 50-74 year-olds were up-to-date with colorectal cancer screening in 2013.[51] Screening compliance varied widely by availability of health care, length of US residence, race, ethnicity and education.[51]

IMPROVING SCREENING PROGRAMS

Any type of screening can be enhanced by reducing the harms and increasing the benefits from that type of screening. Even for cervical cancer screening, which has been ongoing for decades, possibilities for improvement keep arising. The development of new vaccines and tests constantly triggers the evaluation of new prevention strategies. Although colorectal cancer screening is relatively new in most countries, already a wide variety of tests exists, and new tests are constantly being developed. In this thesis, several suggestions for improvement of the harm-benefit ratio of cervical and colorectal cancer screening are described, by either focusing on reducing the harms or increasing the benefits from screening, or on a combination of both.

Reducing harms

It is important to keep track of the potential harms of screening. For cervical cancer screening, switching from cytology-based screening to HPV-based screening may lead to increased harms due to an increased number of (false-)positive screening tests. For women who are currently screened more often than recommended by international guidelines (e.g. the yearly check-ups in Germany), the expected increase in benefits with HPV screening may be outweighed by the increase in burden due to the increase in positive tests and unnecessary follow-up examinations. In **Chapter 3**, the potential harms of primary HPV testing in over-screened women is discussed in further detail.

Another potential harm of cervical cancer screening is the increased risk of preterm birth after loop electrosurgical excision procedure (LEEP), i.e. the most commonly used CIN treatment method.[52,53] As especially at young age many CIN lesions regress naturally, the fact that treatment may be harmful emphasizes the importance of a restrained treatment policy in women of reproductive age. A cost-benefit analysis of screening in reproductive age, taking into account the increased risk of preterm birth after CIN treatment, is presented in **Chapter 4**.

Increasing benefits

Two ways of enhancing the benefits from screening are by increasing participation rates, and by offering a test with better test characteristics.

Over the past 15 years, conventional cytology has been gradually replaced by the LBC tests SurePath and ThinPrep in, amongst others, the Dutch cervical cancer screening program. As said, the use of LBC involves fewer samples of unsatisfactory quality and enables the possibility of also testing samples for the presence of HPV. As LBC test characteristics were thought to be non-inferior to those of conventional cytology,[54,55] these two benefits led to the implementation of LBC in many countries. However, the systematic reviews on LBC versus conventional cytology did not consider a potential difference in test characteristics between SurePath and ThinPrep. Using population-based data, we have shown that SurePath detects more CIN 2+ than conventional cytology, while ThinPrep does not.[56]

This suggests that the sensitivity of SurePath may be higher than that of conventional cytology and ThinPrep. However, the additionally detected CIN lesions might only include non-progressive lesions. Whether SurePath has a higher sensitivity for progressive lesions can only be evaluated by comparing cervical cancer rates. Therefore, in **Chapter 5** of this thesis, we compare the cumulative cancer incidence after a normal conventional cytology, SurePath and ThinPrep screening sample.

In January 2017, the Dutch cervical cancer screening program has changed from cytology-based screening to HPV-based screening. This change is expected to lead to an increased program sensitivity, preventing over 100 additional cervical cancer cases and ~35 additional cervical cancer deaths annually.[57] Another advantage of HPV testing is that it can be performed on self-collected samples. This may trigger women who would not participate in screening otherwise, to take a sample and send it to the laboratory. In **Chapter 6** of this thesis, the cost-effectiveness of offering a self-sampling test to non-attending women is evaluated, considering a range of different scenarios.

For colorectal cancer screening, a variety of different tests is available. The US Preventive Services Task Force included 6 different tests in its 2016 recommendations.[58] The large variety in tests stimulates individuals who do not want to be screened with test X, to still participate in screening with test Y. Although not all tests are equally effective and cost-effective, offering an alternative test with slightly worse test characteristics may increase the expected benefits of screening if it entices previously unscreened individuals to participate in screening. Therefore, the Centers for Medicare and Medicaid Services (CMS) recently started covering the first multitarget Stool DNA (mtSDNA) test, Cologuard®, in the Medicare population. In addition to the ability of regular stool-based tests to identify the presence of occult hemoglobin in stool, the mtSDNA test can identify multiple human DNA biomarkers. The cost-effectiveness of three-yearly mtSDNA testing in the Medicare population, as is covered by CMS, is evaluated in **Chapter 7** of this thesis.

Reducing harms and increasing benefits: Opportunities for risk-based screening

Current screening programs are often "one size fits all", but there is an increasing demand for a more personalized approach. A personalized approach implies that individuals at increased risk are offered more intensive screening, while individuals at decreased risk are offered less intensive screening. By allocating screening resources based on disease risk, screening resources are used more efficiently. In theory, personalizing a screening program therefore always improves its harm-benefit ratio. However, it requires a more advanced organizational structure and the willingness of the screening eligible population to adhere to such guidelines.

For cervical cancer, the future screening eligible population can be divided into two main risk groups: vaccinated and unvaccinated women. Routine vaccination will not only reduce cervical cancer risk with roughly 80% in vaccinated women, but will also reduce HPV prevalence in the general population. This so-called herd immunity is likely to increase over time,

thereby gradually also reducing the cervical cancer risk in unvaccinated women. At first, vaccinated women will likely be at much lower risk of developing cervical cancer than unvaccinated women. However, they may not agree with being offered less intensive screening, just because they were willing to get vaccinated. In **Chapter 8**, we have investigated at what level of herd immunity it would be cost-effective to replace a screening program that is optimized to the risk level in a pre-vaccination cohort with one that is optimized to the risk level in a fully vaccinated cohort.

For colorectal cancer, risk groups can, among other things, be identified based on genetics. For example, someone's family history is indicative of someone's genetic predisposition for the disease. Therefore, several guidelines recommend more intensive screening for those with a family history of colorectal cancer compared to the general population. We previously showed that it is more cost-effective to screen persons with multiple family members affected by the disease more intensely than persons with only one affected family member.[59] In **Chapter 9** of this thesis, we evaluate whether we could further optimize screening for different levels of family history by also taking age of the person at risk into consideration.

We also more directly considered genetic factors for personalized screening, i.e. by considering colorectal cancer screening based on polygenic risk. Polygenic testing can be used to reveal the presence or absence of single nucleotide polymorphisms (SNPs) that are associated with colorectal cancer risk. In **Chapter 10**, we investigated the potential benefit of offering all individuals a polygenic test at a certain age, with future colorectal cancer screening based on the results of that test. We compared the costs and effects of such individualized screening to current uniform screening.

MICROSIMULATION MODELING

MISCAN model

The simulation studies described in this thesis were performed using the microsimulation screening analysis (MISCAN) model, which has been developed at the Department of Public Health of the Erasmus University Medical Center.[60] This model can be used to assess the harms and benefits of different screening programs. Although different versions of the model have been developed for different cancer sites, the overall structure of the model is the same (**Figure 5**). The general idea behind the model is simple; we simulate a hypothetical cohort of individuals who are exposed to the disease. We let the model run twice; once with screening and once without screening. The effects of screening are then determined by the difference in results between those simulations.

Although demography inputs such as life expectancy can easily be derived from population databases, assumptions for screening and natural history parameters might not be readily available. Even though much research has been done on the natural history of diseases, it is impossible to obtain the exact underlying disease process from clinical studies.

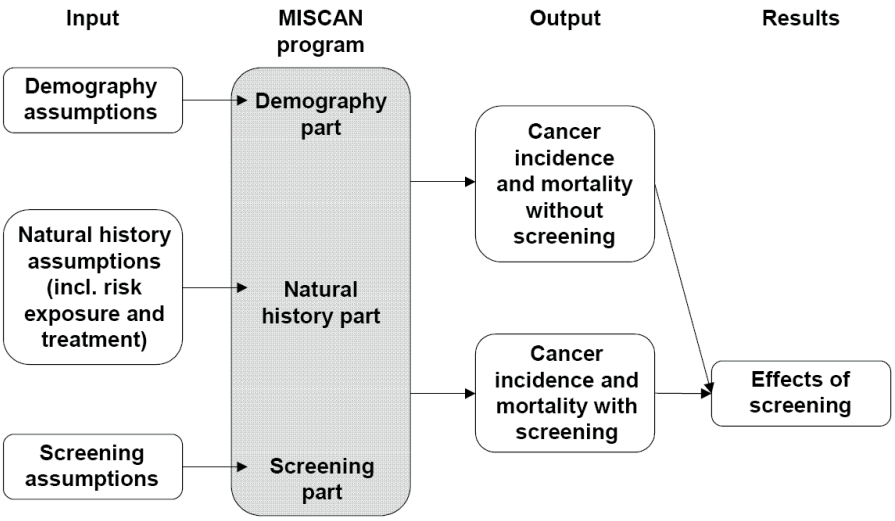


Figure 5. Structure of the MISCAN model.

Model calibration is therefore needed to estimate most of the natural history parameters, such as the duration of preclinical disease stages.

MISCAN is an individual-based model, meaning that individual life histories are simulated one by one. First, a date of birth and a date of death is generated, resulting in a life history without disease (top line **Figure 6**). Second, the model simulates the development of precursor lesions. While most individuals will not develop any of those lesions, some might develop multiple. In this example, the individual develops one precursor lesion, at age 40. In the absence of screening, the precursor lesion will develop to clinical cancer and will lead to death from cancer at age 75. However, if a successful screening intervention takes place at age 50, the precursor lesion is detected and can be removed before it would have progressed to cancer. By avoiding death from cancer, the simulated individual will gain five life years (bottom line **Figure 6**).

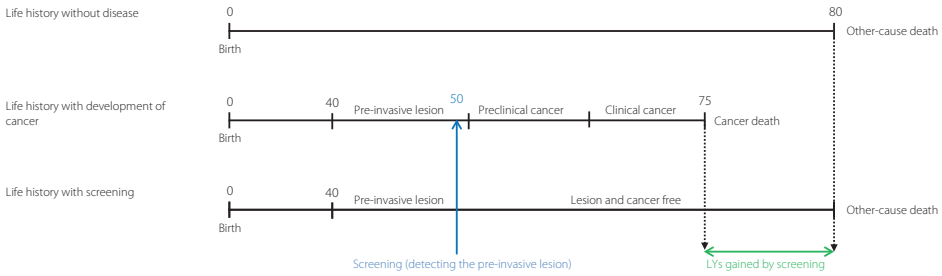


Figure 6. Example of a life history as simulated in MISCAN.

Why using a model helps to answer policy questions

Although RCTs are indispensable, they can only evaluate a few intervention strategies at a time, are expensive, and rely on the willingness of individuals to participate. Furthermore, the effectiveness of a cancer screening strategy in terms of incidence or mortality reduction can only be evaluated after several years of follow-up. In theory, screening strategies can alter in various ways, with differences in screening test, age range, screening interval, and in some instances also the referral threshold for follow-up testing. The options are so numerous, that it is impossible to compare them all in an RCT. This is one of the reasons why mathematical models have been developed for the evaluation of different screening strategies. In those models, any type of screening strategy can be simulated, and results can be generated within a reasonable amount of time (i.e. millions of individuals can be simulated within a few minutes). Models can be adjusted to other settings, and can be used to reproduce the results of trials. In summary, RCTs are needed to prove the effectiveness of screening, but microsimulation models are needed to evaluate a wide range of screening interventions.[61]

RESEARCH QUESTIONS AND OUTLINE OF THIS THESIS

Current literature already includes a wide range of studies on cervical and colorectal cancer screening. The aim of this thesis is to provide insight into some of the gaps in literature on the possibilities to increase the harm-benefit ratio of both types of screening. The remainder of this thesis is subdivided into four parts. Whereas **Part I** elaborates on a methodological issue in cost-effectiveness analyses, the remainder includes studies on how to reduce harms (**Part II**), increase benefits (**Part III**), and improve the harm-benefit ratio by offering risk-stratified screening (**Part IV**).

The research questions addressed in the respective parts are as follows.

Part I. Methodological issues in cost-effectiveness analyses

- To what extent do cost-effectiveness analyses of cervical cancer screening omit relevant strategies, and how does this affect their conclusions? (**Chapter 2**)

Part II. Reducing harms

- What are the potential harms of primary HPV screening in over-screened women? (**Chapter 3**)
- What is the impact of cervical screening on preterm birth? Is it of such importance that cervical screening at reproductive age should depend on a woman's childwish? (**Chapter 4**)

Part III. Increasing benefits

- Does cervical cancer incidence after a normal cytological sample differ between SurePath, ThinPrep and conventional cytology? (**Chapter 5**)
- When do the harms of offering HPV self-sampling to non-attendees of organized primary HPV screening outweigh its benefits? (**Chapter 6**)
- Is colorectal cancer screening with the multitarget stool DNA test a cost-effective alternative for the Medicare population, and if not, under what conditions will it be? (**Chapter 7**)

Part IV. Reducing harms and increasing benefits: Opportunities for risk-based screening

- At what level of herd immunity can uniform cervical cancer screening be adjusted to the risk level in vaccinated women? (**Chapter 8**)
- Should colorectal cancer screening for people with a positive family history vary by age? (**Chapter 9**)
- What is the potential benefit of risk-stratified colorectal cancer screening based on common genetic variants? (**Chapter 10**)

This thesis ends with a general discussion (**Chapter 11**) in which the above research questions are answered, and suggestions for future research are made.

REFERENCES

1. Miles A, Cockburn J, Smith RA, et al. A perspective from countries using organized screening programs. Cancer 2004;101:1201-13.
2. Wilson JM, Jungner YG. [Principles and practice of mass screening for disease] Principios y metodos del examen colectivo para identificar enfermedades. Bol Oficina Sanit Panam 1968;65:281-393.
3. Andermann A, Blancquaert I, Beauchamp S, et al. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. Bull World Health Organ 2008;86:317-9.
4. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66:7-30.
5. International Agency for Research on Cancer. GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012. Volume 2016, 2015.
6. Pimenta JM, Galindo C, Jenkins D, et al. Estimate of the global burden of cervical adenocarcinoma and potential impact of prophylactic human papillomavirus vaccination. BMC Cancer 2013;13:553.
7. Walboomers JM, Jacobs MV, Manos MM, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 1999;189:12-9.
8. McCredie MR, Sharples KJ, Paul C, et al. Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study. Lancet Oncol 2008;9:425-34.
9. Massad LS, Einstein MH, Huh WK, et al. 2012 updated consensus guidelines for the management of abnormal cervical cancer screening tests and cancer precursors. Obstet Gynecol 2013;121:829-46.
10. Public Health England. NHS Cervical Screening Programme: Colposcopy and Programme Management, 2016.
11. Kahn JA, Rosenthal SL, Succop PA, et al. The interval between menarche and age of first sexual intercourse as a risk factor for subsequent HPV infection in adolescent and young adult women. J Pediatr 2002;141:718-23.
12. Burchell AN, Winer RL, de Sanjose S, et al. Chapter 6: Epidemiology and transmission dynamics of genital HPV infection. Vaccine 2006;24 Suppl 3:S3/52-61.
13. Poljak M, Kocjan BJ, Potocnik M, et al. Anogenital hairs are an important reservoir of alpha-papillomaviruses in patients with genital warts. J Infect Dis 2009;199:1270-4.
14. Manhart LE, Koutsky LA. Do condoms prevent genital HPV infection, external genital warts, or cervical neoplasia? A meta-analysis. Sex Transm Dis 2002;29:725-35.
15. Lam JU, Rebolj M, Dugué PA, et al. Condom use in prevention of Human Papillomavirus infections and cervical neoplasia: systematic review of longitudinal studies. J Med Screen 2014;21:38-50.
16. Roura E, Castellsague X, Pawlita M, et al. Smoking as a major risk factor for cervical cancer and precancer: results from the EPIC cohort. Int J Cancer 2014;135:453-66.
17. Luhn P, Walker J, Schiffman M, et al. The role of co-factors in the progression from human papillomavirus infection to cervical cancer. Gynecol Oncol 2013;128:265-70.
18. Gierisch JM, Coeytaux RR, Urrutia RP, et al. Oral contraceptive use and risk of breast, cervical, colorectal, and endometrial cancers: a systematic review. Cancer Epidemiol Biomarkers Prev 2013;22:1931-43.
19. Scarinci IC, Garcia FA, Kobetz E, et al. Cervical cancer prevention: new tools and old barriers. Cancer 2010;116:2531-42.
20. Conti M, Agarossi A, Parazzini F, et al. HPV, HIV infection, and risk of cervical intraepithelial neoplasia in former intravenous drug abusers. Gynecol Oncol 1993;49:344-8.
21. Wright TC, Jr., Ellerbrock TV, Chiasson MA, et al. Cervical intraepithelial neoplasia in women infected with human immunodeficiency virus: prevalence, risk factors, and validity of Papanicolaou smears. New York Cervical Disease Study. Obstet Gynecol 1994;84:591-7.

22. Serraino D, Carrieri P, Pradier C, et al. Risk of invasive cervical cancer among women with, or at risk for, HIV infection. Int J Cancer 1999;82:334-7.
23. Maiman M, Fruchter RG, Guy L, et al. Human immunodeficiency virus infection and invasive cervical carcinoma. Cancer 1993;71:402-6.
24. Thompson D, Easton DF, Breast Cancer Linkage C. Cancer Incidence in BRCA1 mutation carriers. J Natl Cancer Inst 2002;94:1358-65.
25. Bruni L, Diaz M, Barrionuevo-Rosas L, et al. Global estimates of human papillomavirus vaccination coverage by region and income level: a pooled analysis. Lancet Glob Health 2016;4:e453-63.
26. Guan P, Howell-Jones R, Li N, et al. Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer. Int J Cancer 2012;131:2349-59.
27. Lehtinen M, Paavonen J, Wheeler CM, et al. Overall efficacy of HPV-16/18 AS04-adjuvanted vaccine against grade 3 or greater cervical intraepithelial neoplasia: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. Lancet Oncol 2012;13:89-99.
28. Future I/II Study Group, Dillner J, Kjaer SK, et al. Four year efficacy of prophylactic human papillomavirus quadrivalent vaccine against low grade cervical, vulvar, and vaginal intraepithelial neoplasia and anogenital warts: randomised controlled trial. BMJ 2010;341:c3493.
29. Arbyn M, Abarca M. Is liquid based cytology an effective alternative for the conventional pap smear to detect cervical cancer precursors? A systematic review and meta-analysis. Report No. 10. Brussels, Belgium: Scientific Institute of Public Health; 2003.
30. National Institute for Public Health and the Environment. [Screening program will change] Bevolkingsonderzoek gaat veranderen.
31. Australian Government: Department of Public Health. National Cervical Screening Program: Future changes to cervical screening.
32. Medical Services Advisory Committee (MSAC). National Cervical Screening Program Renewal: Evidence Review. In: Ageing DoHa, ed. Canberra, 2013.
33. Haggar FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. Clin Colon Rectal Surg 2009;22:191-7.
34. Huxley RR, Ansary-Moghaddam A, Clifton P, et al. The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence. Int J Cancer 2009;125:171-80.
35. Strum WB. Colorectal Adenomas. N Engl J Med 2016;374:1065-75.
36. Morson B. President's address. The polyp-cancer sequence in the large bowel. Proc R Soc Med 1974;67:451-7.
37. Parkin DM, Boyd L, Walker LC. 16. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. Br J Cancer 2011;105 Suppl 2:S77-81.
38. Fearnhead NS, Wilding JL, Bodmer WF. Genetics of colorectal cancer: hereditary aspects and overview of colorectal tumorigenesis. Br Med Bull 2002;64:27-43.
39. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015;64:1637-49.
40. Mandel JS, Church TR, Ederer F, et al. Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. J Natl Cancer Inst 1999;91:434-7.
41. Scholefield JH, Moss S, Sufi F, et al. Effect of faecal occult blood screening on mortality from colorectal cancer: results from a randomised controlled trial. Gut 2002;50:840-4.
42. Kronborg O, Jorgensen OD, Fenger C, et al. Randomized study of biennial screening with a faecal occult blood test: results after nine screening rounds. Scand J Gastroenterol 2004;39:846-51.

43. Lindholm E, Brevinge H, Haglind E. Survival benefit in a randomized clinical trial of faecal occult blood screening for colorectal cancer. Br J Surg 2008;95:1029-36.
44. Shaukat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for colorectal cancer. N Engl J Med 2013;369:1106-14.
45. Zheng S, Chen K, Liu X, et al. Cluster randomization trial of sequence mass screening for colorectal cancer. Dis Colon Rectum 2003;46:51-8.
46. Thiis-Evensen E, Hoff GS, Sauar J, et al. Population-based surveillance by colonoscopy: effect on the incidence of colorectal cancer. Telemark Polyp Study I. Scand J Gastroenterol 1999;34:414-20.
47. Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. Lancet 2010;375:1624-33.
48. Segnan N, Armaroli P, Bonelli L, et al. Once-only sigmoidoscopy in colorectal cancer screening: follow-up findings of the Italian Randomized Controlled Trial--SCORE. J Natl Cancer Inst 2011;103:1310-22.
49. Schoen RE, Pinsky PF, Weissfeld JL, et al. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. N Engl J Med 2012;366:2345-57.
50. Holme O, Loberg M, Kalager M, et al. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: a randomized clinical trial. JAMA 2014;312:606-15.
51. Smith RA, Andrews K, Brooks D, et al. Cancer screening in the United States, 2016: A review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin 2016;66:96-114.
52. Kyrgiou M, Tsoumpou I, Vrekoussis T, et al. The up-to-date evidence on colposcopy practice and treatment of cervical intraepithelial neoplasia: the Cochrane colposcopy & cervical cytopathology collaborative group (C5 group) approach. Cancer Treat Rev 2006;32:516-23.
53. Kyrgiou M, Athanasiou A, Paraskevaidi M, et al. Adverse obstetric outcomes after local treatment for cervical preinvasive and early invasive disease according to cone depth: systematic review and meta-analysis. BMJ 2016;354:i3633.
54. Arbyn M, Bergeron C, Klinkhamer P, et al. Liquid compared with conventional cervical cytology: a systematic review and meta-analysis. Obstet Gynecol 2008;111:167-77.
55. Whitlock EP, Vesco KK, Eder M, et al. Liquid-based cytology and human papillomavirus testing to screen for cervical cancer: a systematic review for the U.S. Preventive Services Task Force. Ann Intern Med 2011;155:687-97, W214-5.
56. Rozemeijer K, Penning C, Siebers AG, et al. Comparing SurePath, ThinPrep, and conventional cytology as primary test method: SurePath is associated with increased CIN II(+) detection rates. Cancer Causes Control 2016;27:15-25.
57. Naber SK, Matthijsse SM, Jansen EEL, et al. [Effects and costs of the renewed cervical cancer screening program in the Netherlands taking into account recent developments] Effecten en kosten van het vernieuwde bevolkingsonderzoek naar baarmoederhalskanker in Nederland naar aanleiding van recente ontwikkelingen. Rotterdam: Erasmus MC, University Medical Center Rotterdam, 2016.
58. U. S. Preventive Services Task Force, Bibbins-Domingo K, Grossman DC, et al. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. JAMA 2016;315:2564-75.
59. Wilschut JA, Steyerberg EW, van Leerdam ME, et al. How much colonoscopy screening should be recommended to individuals with various degrees of family history of colorectal cancer? Cancer 2011;117:4166-74.
60. Habbema JD, van Oortmarssen GJ, Lubbe JT, et al. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed 1985;20:79-93.
61. van Hees F, Zauber AG, van Veldhuizen H, et al. The value of models in informing resource allocation in colorectal cancer screening: the case of The Netherlands. Gut 2015;64:1985-97.

PART I

Methodological Issues in Cost-Effectiveness Analyses

Chapter 2

Beware of kinked frontiers: a systematic review of the choice of comparator strategies in cost-effectiveness analyses of human papillomavirus testing in cervical screening

James F. O'Mahony, Steffie K. Naber, Charles Normand, Linda Sharp, John J. O'Leary, Inge M.C.M. de Kok

Value in Health. 2015 Dec;18(8):1138-51.

ABSTRACT

Objectives: To systematically review the choice of comparator strategies in cost-effectiveness analyses (CEAs) of human papillomavirus testing in cervical screening.

Methods: The PubMed, Web of Knowledge, and Scopus databases were searched to identify eligible model-based CEAs of cervical screening programs using human papillomavirus testing. The eligible CEAs were reviewed to investigate what screening strategies were chosen for analysis and how this choice might have influenced estimates of the incremental cost-effectiveness ratio (ICER). Selected examples from the reviewed studies are presented to illustrate how the omission of relevant comparators might influence estimates of screening cost-effectiveness.

Results: The search identified 30 eligible CEAs. The omission of relevant comparator strategies appears likely in 21 studies. The ICER estimates in these cases are probably lower than would be estimated had more comparators been included. Five of the 30 studies restricted relevant comparator strategies to sensitivity analyses or other sub-analyses not part of the principal base-case analysis. Such exclusion of relevant strategies from the base-case analysis can result in cost-ineffective strategies being identified as cost-effective.

Conclusions: Many of the CEAs reviewed appear to include insufficient comparator strategies. In particular, they omit strategies with relatively long screening intervals. Omitting relevant comparators matters particularly if it leads to the underestimation of ICERs for strategies around the cost-effectiveness threshold because these strategies are the most policy relevant from the CEA perspective. Consequently, such CEAs may not be providing the best possible policy guidance and lead to the mistaken adoption of cost-ineffective screening strategies.

INTRODUCTION

This review considers the choice of screening strategies compared in cost-effectiveness analyses (CEAs) of cancer screening programs. It investigates how the choice of which strategies are compared can influence cost-effectiveness estimates and resulting policy advice. Specifically, this review addresses the choice of comparator strategies against which the cost-effectiveness of a given screening strategy is estimated. This issue is considered in the particular context of CEAs of cervical cancer screening using testing for the human papillomavirus (HPV).

The primary measure of cost-effectiveness is the incremental cost-effectiveness ratio (ICER), which is the ratio of additional costs to additional health effects of an intervention relative to its next best alternative (once strategies subject to simple and extended dominance have been eliminated).[1,2] Because the ICER is an incremental measure, it depends not only on the costs and effects of the strategy for which it is estimated but also on those of the comparator strategy.

Typically, decision makers use ICERs in conjunction with a cost-effectiveness threshold, which indicates the maximum willingness to pay for an additional (quality-adjusted) life year.[3] The strategy with the highest ICER within the threshold is optimal from the cost-effectiveness perspective because it is the most effective intervention that does not exceed the willingness-to-pay limit. More broadly, those strategies with ICERs closest to the threshold are defined here as the CEA-relevant strategies because they yield more net health benefit than do strategies with ICERs far above or below the threshold. It is the adequacy of the choice of comparators for these CEA-relevant strategies that is the focus of this review.

A particular characteristic of screening especially relevant to CEA modeling is that it can often be applied at a wide range of intensities, depending on the screening interval, screening age range, type of tests used, and the diagnostic criteria for follow-up. As a result, CEAs considering a wide range of screening intensities can yield a wide range of ICERs, varying from those well below the threshold through to those around the threshold and then on to well above the threshold.

What is already well appreciated in CEA theory is the importance of including relevant comparators for the reliable estimation of ICERs. Indeed, the Washington Panel on Cost-Effectiveness in Health and Medicine uses a cancer screening example to illustrate the importance of including relevant comparators.[4] It notes that to correctly estimate the ICER of annual screening, it must be compared with biennial screening rather than with no screening. The general principle is that to appraise the cost-effectiveness of a given screening strategy, the next best strategy should be included as a comparator against which to estimate the ICER. If less intense comparators are omitted, then the estimated ICER is likely to be lower than that in a more complete comparison, thereby giving an unrepresentatively favorable impression of the strategy's cost-effectiveness.

The motivation for this review was an observation that although most models used in CEAs of HPV screening are carefully constructed and well described, many include relatively

few comparator screening strategies. Consequently, they may fail to adequately estimate the cost-effectiveness of certain strategies. This, in turn, could lead decision makers to mistakenly adopt cost-ineffective policies, thereby wasting health care resources. Therefore, the aim of this study was to systematically assess the adequacy of the choice of comparator strategies in CEAs of HPV testing in cervical screening. It seeks to demonstrate the importance of appropriately chosen comparators for the reliable estimation of ICERs. Although the review addresses the specific case of cervical screening, it is hoped that the example will illustrate the importance of including relevant comparators in CEAs in general to both analysts and decision makers alike.

The Example of Cervical Screening

Cervical screening has proved highly successful in reducing cancer incidence and mortality.[5] Cervical screening is widely practiced in developed countries, either through organized programs or on an ad-hoc basis.[6,7] There is a wide variety of possible screening strategies because alternative screening intervals and start and stop ages can be used. Similarly, screening may use different tests, such as conventional Papanicolaou cytology or the more recent alternative of liquid-based cytology. Furthermore, there are alternative combinations of primary screening tests and triage testing for inconclusive primary screen results and alternative classifications of borderline results. In practice, there are large variations between countries in screening recommendations. For example, the German recommendations are for annual screening from age 20 years, whereas the Dutch screening program has used screening every 5 years from age 30 years.[8,9]

The range of possible strategies continues to expand, in part because of the recent advent of HPV DNA testing. HPV testing offers better sensitivity for the detection of high-grade lesions, but at the cost of lower specificity.[10–12] HPV testing is typically used in conjunction with cytology, for example, using HPV and cytology as the primary test and the triage test, respectively. Some proposed strategies also involve a switch in the order the strategies are used,[8] using cytology and HPV as the primary test and the triage test, respectively, in younger women in whom transient HPV infections are more prevalent. Importantly, for this review, HPV testing has been recognized as offering the potential for longer screening because a negative HPV test result is associated with a longer period of reduced risk of precancerous lesions than is negative cytology.[13]

Another relevant development is the HPV vaccine, which has been implemented in many countries recently. Although current vaccines are expected to reduce the incidence of cervical cancer, the level of protection is not anticipated to be sufficient to abandon screening.[14] Reduced incidence will reduce the cost-effectiveness of current screening services, so screening intervals may need to lengthen for screening to be cost-effective.[15]

METHODS

The PubMed, Web of Knowledge, and Scopus databases were searched for model-based CEAs of cervical screening using HPV testing. The search string is given in **Box 1**. **Figure 1** shows the search protocol. The search was restricted to English language academic articles published between January 1995 and September 2013. The search excluded conference proceedings, government reports, and gray literature. The search returned 382, 438, and 361 titles from PubMed, Web of Knowledge, and Scopus, respectively. Combining and removing duplicates gave 646 unique studies, the titles and abstracts of which were then reviewed by one reviewer (J.F.O'M.).

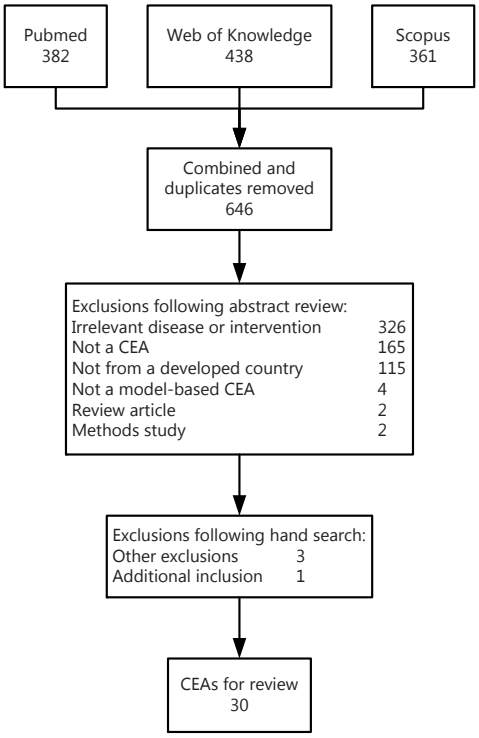


Figure 1. Composition of literature search and exclusions.
CEA = cost-effectiveness analysis.

Box 1. The PubMed version of the search string

Cervi*[tiab] OR pap[tiab] OR cytolog*[tiab] OR (cervi*[tiab] AND cancer[tiab])
AND (HPV[tiab] OR "Human Papillomavirus"[tiab])
AND (screen*[tiab] OR prevent*[tiab])
AND (cost-effect*[tiab] OR "cost effect*"[tiab] OR CEA[tiab] OR CUA[tiab] OR HTA[tiab]OR "health technology assessment"[tiab] OR "health economic"[tiab])
AND English[lang]
AND ("1995/01/01"[PDAT] : "2013/10/01"[PDAT])

Studies were excluded if they did not relate to cervical cancer or cervical screening. Studies were included only if they concerned screening in countries classified as advanced economies by the International Monetary Fund.[16] These studies were excluded for the two reasons that the screening services in such countries are very different from those in developed settings and it is markedly less clear what the appropriate cost-effectiveness thresholds are. Studies were excluded if they were trial-based CEAs because the intermediary outcome measures such as cases detected are not directly comparable with estimates of life years gained (LYG) or quality-adjusted life years (QALYs) from model-based CEAs.[17–21] Methodological studies and reviews were also excluded.[22–25]

The remaining studies were reviewed independently in detail by two reviewers (J.F.O'M. and S.K.N.). Discordance between reviewers regarding the inclusion of articles was resolved by reviewing the studies together. One analysis was published in two similar articles, one in an academic journal,[26] the other in a government journal;[27] the latter was excluded. In addition, the article by van Ballegooijen et al.[28] was excluded because it is primarily an exploratory analysis of two alternative hypotheses regarding disease progression rather than a standard CEA. Another study was excluded because it did not report costs and effects estimates or a cost-effectiveness plane, but reported only ICERs, meaning that it was not possible to appraise the appropriateness of the comparisons made.[29] The reference lists of the included studies were reviewed, and this yielded an additional article.[30] The final number of CEAs reviewed was 30.[8,14,15,26,30–55]

The reported costs and effects estimates from the 30 studies were extracted by the two reviewers and compiled for analysis. The cost-effectiveness planes and ICERs were reproduced from the available results. Some studies reported ICERs that are at variance with the conventional interpretation as the ratio of incremental costs to incremental health effects relative to the next most effective strategy.[38,52,54,55] In these cases, the ICERs were recalculated from the reported costs and effects. The reproduced cost-effectiveness planes were used to review the comparisons and interpretations made by the studies.

The selected studies were reviewed to assess the adequacy of the comparators included. This was informed by a previous CEA of cervical screening that assessed conventional cytology over a broad range of screening ages and intervals.[56] It estimated an efficient frontier without pronounced kinks, as the ICERs increased steadily from long intervals and short screening age ranges (low-intensity strategies) through to very high ICERs for short screening intervals and long screening ages ranges (high-intensity strategies). It was assumed that if the reviewed studies also simulated a broad range of screening intensities they would find similarly shaped frontiers, in which low-intensity strategies would be the comparators against which the ICERs of higher-intensity strategies were estimated. This assumption implies that if low-intensity comparators have been omitted from the analysis, then some of the resulting ICER is likely to be underestimated.

The primary objective was to assess the adequacy of the comparator against which the ICER of the optimal strategy was estimated in each study. The cost-effectiveness threshold used to identify the optimal strategy was retrieved from each study, or, if not explicitly stated,

from another included study for the same country. In one case this was not possible,[26] so the gross domestic product per capita threshold suggested by the World Health Organization from the year of publication was used instead.[57,58]

The primary criterion used to identify the omission of a relevant comparator was the failure to include a screening interval longer than that found for the optimal strategy. Studies without such a longer interval strategy were considered likely to have omitted relevant comparators. The analysis also considered the ratio of the ICER of the optimal intervention relative to the next higher ICER on the efficient frontier. Larger ratios were considered more suggestive of the omission of relevant comparators. Finally, the analysis also considered other aspects of the strategies compared, including the screening age ranges, the alternative screening technologies, and the range of primary screening and triage protocols simulated.

This review is based on the assumption that all possible strategies are potential comparators. Some CEAs, however, may be conducted with a specific research objective that (implicitly or explicitly) precludes some comparators. For example, the objective of Vijayaraghavan et al.[54] to assess the cost-effectiveness of adding HPV triage testing to current screening services in the United States may preclude alternative screening intervals. Despite this, the criteria described above were applied to all studies irrespective of the stated research objectives.

Some studies assessed more than one population; simulating either vaccinated and unvaccinated women or a number of countries or provinces. In such cases, the issue of comparator omission was sometimes apparent in some but not all populations. This analysis reports results for subgroups in which the issue of comparator omission is suspected.

The results are further illustrated by considering cost-effectiveness planes for four examples from the review. In three of these examples, additional points marking hypothetical costs and effects estimates for longer screening intervals show how the efficient frontier might change if additional comparators were included. The positions of these hypothetical comparators are informed by inference from other studies and the assumption of decreasing marginal returns in terms of costs and effects as the screening interval is shortened.

RESULTS

First, an overview of the studies included in this review is presented, summarizing the simulated populations and strategies. The principal results of this review are then shown, detailing which studies have likely omitted relevant comparators. These results are then complemented with several specific examples to illustrate the issues in greater detail.

Overview of Eligible CEAs

Details of the eligible studies are presented in **Table 1**. The 30 CEAs were published between 2002 and 2013. The most frequent countries of origin were the United States and The Netherlands with eight studies each, followed by Canada with five.

Table 1. Summary of key attributes and strategies compared in reviewed studies.

Study	Country	Number of strategies modeled	Alternative screening intervals, years	Screening age ranges		Screening switch ages[a]	Vaccinated population	Choice of modeled strategies justified
				Start age	Stop age			
Accetta et al. (2010)	Italy	18	3,5	25	65	-	Yes	No
Balasubramanian et al. (2010)	US	15	1,2,3	18	85	-	No	No
Berkhof et al. (2006)	The Netherlands	20	3 (5 in S.A.)	30	60	-	No	No
Berkhof et al. (2010)	The Netherlands	13	5,6,7.5,10	30	60	-	No	No
Bidus et al. (2006)	US	10	1,2,3	18	85	30	No	No
Bistoletti et al. (2008)	Sweden	4	3,4,5,9	32	60	-	No	No
Burger et al. (2012)	Norway	98	3,4,5,6	25	70	31, 34	Yes	Yes
Chen et al. (2011)	Taiwan	5	1,3,5	30	70	-	Yes	No
Chow et al. (2010)	Taiwan	10	1,3,5	30	69	-	No	No
Chuck (2010)	Canada	21	1,2,3	18	69	30	No	No
Coupé et al. (2009)	The Netherlands	10	5,6,7.5,10	30-35	60	-	Yes	No
Coupé et al. (2012)	The Netherlands	10	5,6,7.5,10	30	60	-	Yes	No
Diaz et al. (2010)	Spain	333[b]	1,2,3,4,5	18,25,30	50,65,85	35, 40	Yes	No
Goldhaber-Fiebert et al. (2008)	US	186	1,2,3,5	18,21,25	N.S.	25,30,35	Yes	No
Goldie et al. (2004)	US	17	1,2,3,4	18	N.S.	30	No	No
Kim et al. (2002)	US	57	1,2,3,5	18	N.S.	-	No	No
Kim et al. (2005)	UK, The Netherlands, France, Italy	6-8	3,5	25,30	60,65	30	No	No
de Kok et al. (2012)	The Netherlands	1,539	3,4,5,6, 7,8,9,10	25,27, 30,32	31-70	-	Yes	No
Kulasingam et al. (2009)	Canada	27	1,2,3,5	18,25	70	-	No	No

Table 1 (continued.) Summary of key attributes and strategies compared in reviewed studies.

Study	Country	Number of strategies modeled	Alternative screening intervals, years	Screening age ranges		Screening switch ages[a]	Vaccinated population	Choice of modeled strategies justified
				Start age	Stop age			
Legood et al. (2006)	UK	5	-	25	64	35	No	No
Mandelblatt et al. (2002)	US	19	2,3	20	65,75, lifetime	-	No	Yes
Maxwell et al. (2002)	US	10	1,2,3	18	85	-	No	No
Mittendorf et al. (2003)	Germany	4	-	20	-	-	No	No
Östensson et al. (2010)	Sweden	8	2,3, 3-5,5	23	60 (65 in S.A.)	35	No	No
Rogoza et al. (2008)	Canada, US, The Netherlands, UK, Taiwan	4-6	1,3,5,7	15,18,20, 25,30	52,58,60, 68-70, 85,87, 89,99,100	-	Yes	No
van Rosmalen et al. (2012)	The Netherlands	1,539	3,4,5,6, 7,8,9,10	25,27,30, 32	31-70	33	No	No
Sherlaw-Johnson & Philips (2004)	UK	15	3,5	21	64	30	No	No
Sroczynski et al. (2011)	Germany	18	1,2,3,5	20-25 in S.A.	lifetime	30	No	No
Vijayaraghavan et al. (2010) [c]	Canada	8	1,3	30	N.S.	-	No	No
Vijayaraghavan et al. (2010) [d]	Canada	7	2,3	30	N.S.	-	No	No

S.A. = sensitivity analysis

N.S. = not stated

[a] The screening switch age is the age at which HPV testing is used at the primary screen test and cytology as the triage test, with the converse order being used in younger women

[b] Personal communication

[c] Vijayaraghavan at al. (2010) Cost-effectiveness of High-risk Human Papillomavirus Testing for Cervical Cancer Screening in Québec, Canada; Can J Public Health

[d] Vijayaraghavan at al. (2010) Cost-effectiveness of using human papillomavirus 16/18 genotype triage in cervical cancer screening. Gynecologic Oncology

The number of alternative screening strategies compared varies widely, ranging from 4 to more than 1500. Four studies considered more than 100 strategies,[8,14,42,46] and 21 studies considered fewer than 20 strategies.[15,18,26,30–33,35,36,38,39,41,43,45,48–50,52–55]

The intensity of screening varied widely. Of the 30 CEAs reviewed, all but 2 varied the screening interval.[48,50] Fourteen studies considered only two or three alternative screening intervals.[18,30–32,34,35,38–40,45,49,53–55] Furthermore, only eight studies considered intervals longer than 5 years, all of which were in European settings.[8,15,33,36,37,41,46,52] Similarly, only 10 varied the screening age range.[8,14,15,26,30,41,42,46,47,52] Some considered only small changes to the screening age range, whereas others considered much broader ranges of alternative start and stop ages. Among the studies with fixed screening age range, start ages between 18 and 30 years were typical, as were stop ages between 60 and 70 years.

Primary HPV screening was considered by all but four studies.[34,44,48,52] Alternative triage strategies for borderline primary cytology or positive primary HPV test results were considered by 19 studies.[8,14,31,32,34,35,40–48,52–55] Strategies featuring a switch from primary cytology screening with HPV triage to primary HPV with cytology triage partway through the screening program were considered by 12 studies.[8,14,18,26,35,37, 40,42,43,45,48,53]

The review also recorded which studies explained the rationale for the range of strategies compared, in particular the screening interval and the screening age range. Only two provided explicit justifications for the intervals considered.[30,37] Both explained that the intervals compared were those of policy recommendations, although neither included any additional comparators with longer intervals to serve as comparators to the recommended strategies. Although the remaining studies did not explain the choice of strategies, many mentioned current guidelines, the status quo strategy, or the possibility of lengthening screening intervals with HPV testing when describing what strategies were compared.[8,14,15,26,31, 33–36,38–41,45–55]

Evidence of Omission of Relevant Comparators

Table 2 presents the principal results of the review. It records the threshold and the associated optimal strategy including the optimal interval. Relevant comparators have likely been omitted in 18 of the 30 studies reviewed. These are recorded in the table as having "probably" omitting relevant comparators. In most cases, this conclusion is based on the observation that a longer screening interval than that found to be optimal was not included.

There are three exceptions to the general observation that studies without longer intervals have omitted relevant comparators. Although Rogoza et al.[52] does include longer intervals, the choice of screening age ranges is inconsistent between the alternative intervals simulated. In particular, the annual interval found to be optimal has a relatively short screening age range of 25 to 60 years, whereas the triennial comparator modeled has an age range of 15 to 87 years. The inclusion of a triennial strategy with an age range closer to 25 to 60 years would probably be relevant, as would a biennial comparator. Similarly, the choice of screening protocols is inconsistent between the intervals considered in Chen et al.[38] because combined cytology and HPV testing is not simulated in all intervals considered. Furthermore, although annual cytology is found cost-effective relative to triennial cytology,

no biennial strategy is modeled; despite this, biennial screening is typically found to be a relevant comparator to annual screening in other studies. Finally, although Goldhaber-Fiebert et al.[42] do not include intervals longer than that identified as optimal for vaccinated women, the strategies found to be efficient at shorter intervals suggest that the broad range of screening age ranges and screening protocols simulated probably provide sufficient comparators to correctly estimate the ICER of the optimal strategy.

The likely omission of relevant comparators is quite distinct in some cases, such as Accetta et al.[31] and Bidus et al.[35] This is evidenced in part by a high ratio of the next efficient ICER to that of the optimal strategy. There are other cases in which the omission of comparators is much less certain. For example, although Burger et al.[37], Kim et al.[45], and Sherlaw-Johnson et al.[53] all omit intervals longer than that found to be optimal, the ratio of adjacent ICERs in these cases is much lower and it is less certain whether the ICER of the optimal interval would change with the inclusion of additional comparators.

Illustrative Examples of Omissions

Four examples are now presented to further illustrate the omission of relevant comparators. The first is one of the distinct examples of comparator omission provided by Accetta et al.[31] It assessed four combinations of primary and triage testing at screening intervals of 3 and 5 years. Note that although the original analysis considered vaccinated and vaccinated women together, we consider them separately. **Figure 2A** shows the estimates for unvaccinated women. There is a notably sharp kink in the frontier around the second efficient strategy as the ICER increases from €5,700/QALY to €68,400/QALY. This corresponds to the highest ratio of adjacent ICERs observed in this review of 11.8. Similarly kinked frontiers are also found in other analyses with few alternative intervals.[32,49,53] Such kinked frontiers contrast with the gently curved frontiers found in analyses with a broader range of screening strategies.[8,42]

It is possible to anticipate how the frontier from Accetta et al.[31] might appear if additional strategies with longer intervals of 6, 7, and 8 years were modeled. **Figure 2B** includes three markers showing hypothetical costs and effects estimates for such intervals. The frontier over these supposed points is marked with the dotted line. In such a case, the frontier would not be sharply kinked, but be more gently curved. Furthermore, it can reasonably be assumed that the ICER estimate of €5,700/QALY for the 5-year interval would be revised upward, meaning that the reported ICER is probably an underestimate.

The lack of sufficient comparators in Accetta et al.[31] is easily identified from the noticeably kinked frontier. The same problem of insufficient comparators, however, can also be found in studies considering a broader range of strategies without obviously kinked frontiers. For example, Goldie et al.[43] considers four combinations of primary and triage testing at intervals of 1, 2, 3, and 4 years. The corresponding cost-effectiveness plane in **Figure 3** has been rescaled for clarity, and the points representing no screening and the four annual screening strategies are not shown. Although the frontier appears gently curved, the figure shows that it is certainly possible that a 5-year interval strategy could lie to the southwest of the 4-year

Table 2. Assessment of reviewed studies for omission of relevant comparator strategies.

Study	Example Subgroup	Threshold Stated Explicitly	Threshold	Optimal Strategy
Accetta et al. (2010)	Unvaccinated	Yes	€50,000/QALY	HPV with cytology triage every 5 years
Balasubramanian et al. (2010)	-	No	$50,000/QALY	HPV with cytology triage every 3 years
Berkhof et al. (2006)	-	No	€20,000/QALY	Cytology with combined triage of cytology & HPV at 6 months, every 5 years
Berkhof et al. (2010)	-	Yes	€20,000/QALY	HPV with cytology triage every 5 years
Bidus et al. (2006)	-	Yes	$50,000/LYG	Cytology with HPV triage every 3 years
Bistoletti et al. (2008)	-	No	NA[a]	Combined cytology and HPV testing every 9 years
Burger et al. (2012)	Vaccinated	Yes	$83,000/LYG	HPV with cytology triage every 6 years
Chen et al. (2011)	-	Yes	$40,000/LYG	Cytology alone annually
Chow et al. (2010)	-	Yes	$1,620,000/QALY[e]	HPV with cytology triage every 3 years
Chuck (2010)	-	Yes	$50,000/QALY[d]	Cytology with cytology triage with HPV testing for women over 30 with ASCUS every 3 years
Coupé et al. (2009)	-	Yes	€20,000/QALY	HPV with cytology triage every 7.5 years
Coupé et al. (2012)	-	Yes	€20,000/QALY	HPV with cytology triage every 10 years
Diaz et al. (2010)	Vaccinated	Yes	€30,000/QALY	Cytology with HPV triage every 5 years
Goldhaber-Fiebert et al. (2008)	Vaccinated	Yes	$50,000-$100,000/QALY	Cytology with HPV triage switching to HPV with cytology triage at age 35 every 5 years
Goldie et al. (2004)	-	Yes	$50,000/QALY	Cytology with HPV triage every 4 years
Kim et al. (2002)	-	No	$50,000/LYG	Cytology with reflex HPV triage every 5 years
Kim et al. (2005)	UK	Yes	$30,200/LYG	HPV and cytology combined testing every 5 years
de Kok et al. (2012)	-	Yes	€20,000/QALY	HPV with cytology triage every 6 years[a]

Comparator	Interval longer than optimal interval simulated	Optimal Strategy ICER	Next Efficient Strategy ICER	ICER ratio between optimal and next efficient strategy	Relevant Comparators Omitted
Cytology with HPV triage every 5 years	No	€5,800/ QALY	€68,400/ QALY	11.8	Probably
No screening	No	$9,900/ QALY	$70,200/ QALY	7.1	Probably
Cytology with HPV triage followed by combined cytology & HPV testing at 6 & 18 months, every 5 years	No	€8,700/ QALY	€22,000/ QALY	2.5	Probably
HPV with cytology triage every 6 years	Yes	€18,800/ QALY	€67,100/ QALY	3.6	Probably not
No screening	No	$5,100/ LYG	$56,700/ LYG	11.1	Probably
NA[a]	No	Cost saving	NA	-	Probably
No screening, vaccination only	No	$80,000/ LYG	$92,000/ LYG	1.2	Probably not
Cytology alone every 3 years	Yes	$31,700/ LYG	NA		Probably
HPV with cytology triage every 5 years	Yes	$1,357,700/ QALY[e]	$3,891,300/ QALY[e]	3.1	Probably not
Cytology with cytology triage annually	No	Cost saving	$58,500/ QALY[d]	-	Probably
HPV with cytology triage every 6 years	Yes	€11,100/ QALY	€26,700/ QALY	2.4	Probably not
Cytology with cytology triage every 10 years	No	€6,700/ QALY	€22,300/ QALY	3.3	Probably
No screening		€24,400/ QALY	€97,000/ QALY	4.0	Probably
Cytology with HPV triage switching to HPV every 5 years	No	$41,100/ QALY	$126,100/ QALY	3.1	Probably not
Cytology with cytology triage every 4 years	No	$20,600/ QALY	$95,300/ QALY	4.6	Probably
Cytology with co-collected HPV triage every 5 years	No	$20,300/ LYG	$59,600/ LYG	2.9	Probably
Cytology with HPV triage every 5 years	No	$13,800/ LYG	$33,200/ LYG	2.4	Probably not
HPV with cytology triage every 7 years[a]	Yes	€10,300/ QALY[a]	€21,130/ QALY[a]	2.1	Probably not

Table 2 (continued). Assessment of reviewed studies for omission of relevant comparator strategies.

Study	Example Subgroup	Threshold Stated Explicitly	Threshold	Optimal Strategy
Kulasingham et al. (2006)	Canada	No	$50,000/ QALY[d]	HPV with cytology triage every 3 years from age 18
Legood et al. (2006)	-	No	£20,000-30,000/ QALY	Cytology with combined triage of cytology & HPV & combined surveillance of cytology & HPV every 3 to 5 years
Mandelblatt et al. (2002)	-	No	$50,000/QALY	Cytology alone every 2 years
Maxwell et al. (2002)	-	Yes	$50,000/QALY	Cytology with HPV triage every 3 years
Mittendorf et al. (2003)	-	No	NA	Combined cytology and HPV annually
Östensson et al. (2010)	-	Yes	€80,000/LYG	HPV with HPV triage every 5 years
Rogoza et al. (2008)	US	Yes	$50,000-$100,000/ QALY	Cytology with cytology triage annually between ages 30-60
van Rosmalen et al. (2012)	-	Yes	€20,000/QALY	HPV with cytology triage every 6 years
Sherlaw-Johnson et al. (2004)	-	No	£20,000-30,000/ QALY	Combined cytology & HPV with combined cytology & HPV triage every 5 years
Sroczynski et al. (2011)	-	No	€33,000/QALY	HPV alone every 2 years
Vijayaraghavan et al. (2010)[b]	-	No	$50,000/ QALY[d]	HPV alone every 3 years
Vijayaraghavan et al. (2010) [c]	-	No	$50,000/QALY	Combined cytology & HPV with triage of HPV genotyping every 3 years

Unless marked otherwise currencies are Euro, UK Pounds or US Dollars.

[a] Personal Communication

[b] Vijayaraghavan at al (2010) Cost-effectiveness of High-risk Human Papillomavirus Testing for Cervical Cancer Screening in Québec, Canada; Can J Public Health

[c] Vijayaraghavan at al (2010) Cost-effectiveness of using human papillomavirus 16/18 genotype triage in cervical cancer screening. Gynecologic Oncology

[d] Canadian Dollars

[e] New Taiwan Dollar

Comparator	Interval longer than optimal interval simulated	Optimal Strategy ICER	Next Efficient Strategy ICER	ICER ratio between optimal and next efficient strategy	Relevant Comparators Omitted
HPV with cytology triage every 3 years from age 25	Yes	$47,300/ QALY[d]	$72,000/ QALY[d]	1.5	Probably not
Cytology with combined triage of cytology & HPV & surveillance of cytology alone every 3 to 5 years	No	£18,600 /LYG	NA	-	Probably
Cytology alone every 3 years	Yes	$29,800/ QALY	$56,400 /QALY	1.9	Probably not
Cytology alone every 3 years	No	$14,300/ LYG	$65,500/ LYG	4.6	Probably
HPV alone annually	No	€298/ LYG	NA	-	Probably
No screening	No	€43,000/ LYG	€84,000/ LYG	2.0	Probably
Cytology with cytology triage every 5years between ages 15 and 85	Yes	$47,700/ QALY	$178,100/ QALY	3.7	Probably
HPV with cytology triage every 7 years	Yes	€10,300/ QALY	€21,130/ QALY	2.1	Probably not
HPV with cytology triage every 5 years	No	£22,600/ LYG	£37,900/ LYG	1.7	Probably not
HPV alone every 3 years	Yes	€28,400/ LYG	€93,700/ LYG	3.3	Probably not
Cytology alone every 3 years	No	$11,400/ QALY[d]	NA	-	Probably
Combined cytology & HPV every 3 years	No	$33,500/ QALY	NA	-	Probably

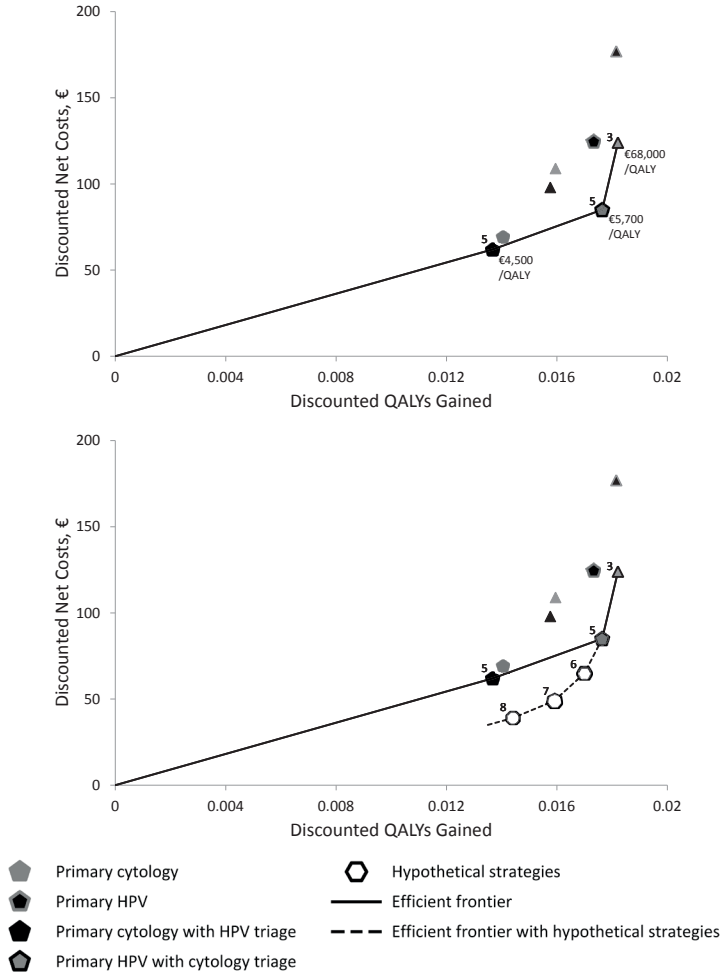


Figure 2A & B. Cost-effectiveness plane for unvaccinated women reinterpreted from Accetta et al. with the triangle and pentagons marking 3 and 5 year interval strategies respectively and panel B including hypothetical estimates for longer intervals of 6, 7 and 8 years and the possible change in the efficiency frontier.

QALY = quality-adjusted life year.

interval strategy, which has an ICER of $20,600/QALY. If a 5-year interval strategy lay at the point marked with the pentagon, this would result in a higher ICER than $20,600/QALY for the 4-year interval strategy.

The omission of 5-year interval comparators in Goldie et al.[43] is particularly relevant because the omitted strategies would likely form the most CEA-relevant part of the efficient frontier. As Goldie et al.[43] note, although there is no established threshold in the United States, a value of $50,000/QALY is commonly used. The strategies with ICERs of $95,300/QALY and above are not CEA-relevant because they exceed the threshold, whereas the

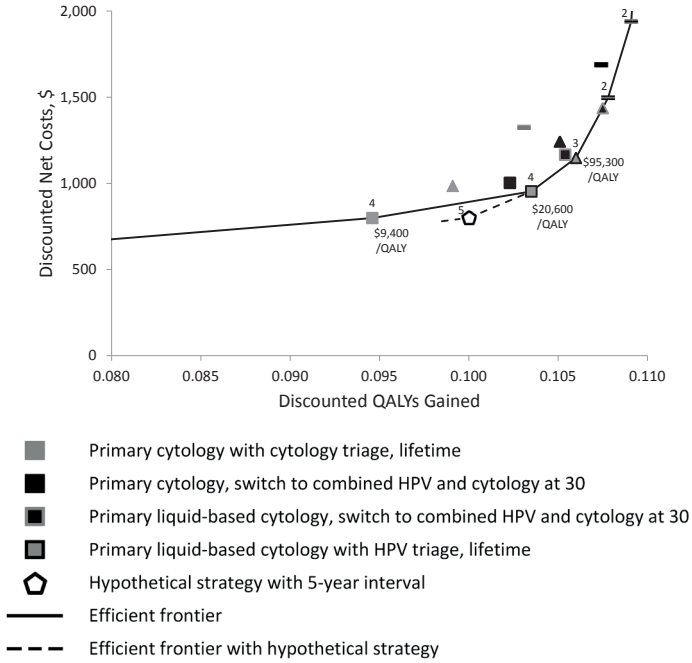


Figure 3. Cost-effectiveness plane from Goldie et al. showing the policy-relevant section of the efficient frontier with various screening strategies with dashes, triangles and squares representing intervals of 2, 3 and 4 years respectively and the inclusion of a hypothetical strategy with an interval of 5 years marked with a pentagon.
QALY = quality-adjusted life year.

$20,600/QALY strategy is optimal. Therefore, it is significant that the inclusion of a 5-year interval comparator would probably inflate this strategy's ICER.

It is not possible to know how much greater the ICER of a 4-year strategy would be if a 5-year interval was included. Had 4-year screening intervals been omitted, however, the ICER of the efficient 3-yearly strategy would be $23,500/QALY rather than $95,300/QALY. This large difference indicates how important the omission of comparators can be.

The importance of the CEA-relevant portion of the frontier is further illustrated by contrasting Goldie et al.[43] with an example in which additional comparators would change the frontier, but not within the CEA-relevant range. **Figure 4** graphs estimates for combinations of cytology and HPV testing at 2- and 3-year intervals from Mandelblatt et al.[30] The 2-year strategy is optimal because it has the highest ICER within the threshold. The inclusion of a 4-year interval comparator, represented here as a hypothetical estimate marked with the square, would probably increase the ICER of 3-yearly screening but not affect that of 2-yearly screening. So although the inclusion of additional comparators with longer intervals would likely change the frontier, the change is below the CEA-relevant portion and is therefore of little significance.

The influence of the choice of comparator strategies can be further illustrated by showing how an analysis simulating a large range of screening alternatives differs when restricted to relatively in gray line in **Figure 5** show the costs and effects estimates and efficient frontier for all strategies assessed in van Rosmalen et al.[8] The costs and effects estimates for the dominated strategies were sourced through personal communication. The estimates have been rescaled to per-woman estimates from the original source for consistency with the other figures presented. This analysis includes a much wider range of screening intervals and age ranges than considered in most of the reviewed studies. The ICERs are included at three points on the frontier to illustrate the broad range of ratios estimated in this case. The analysis can be constrained to higher-intensity strategies typical of many of the studies included in this review. The black points and the black line show the estimates and frontier for strategies with intervals no longer than 5 years and screening age ranges that start no later and finish no earlier than ages 25 and 60 years, respectively.

Such a restriction omits very many of the efficient strategies from the complete analysis and finds strategies that are dominated in the complete analysis to be efficient. The optimal strategies given the Dutch threshold of €20,000/QALY in the complete and restricted analyses are circled. The restricted analysis would lead to the selection of a program with eight rather than three lifetime screens that costs approximately 2.5 times more. Finally, the only efficient strategy common to both analyses is the most costly and most effective strategy. The restricted analysis estimates an ICER of €68,700/QALY for this strategy, whereas the unrestricted analysis finds a considerably higher ICER of €122,500/QALY.

Excluding Relevant Comparators from the Base Case

A secondary issue with the choice of comparators identified by this review is the exclusion of strategies from the base case that are simulated elsewhere in the analysis. This issue occurs in 5 of the 30 studies reviewed. Three studies featured additional strategies in secondary analyses that were not included in the base case.[8,26,37] Importantly, some of these additional strategies dominate strategies found to be efficient in the base case. For example, **Figure 5** shows the cost-effectiveness estimates from van Rosmalen et al.[8], including two strategies marked with the hollow squares considered in a sensitivity analysis in which the order of cytology and HPV testing between primary screening and triage is reversed at age 33 years.

Two other studies estimated costs and effects for a range of screening strategies, but excluded strategies that were less effective than the status quo when estimating ICERs, instead using the status quo as the comparator.[33,40] Consequently, in both studies, the ICER estimate for one strategy is underestimated and the ICER estimates for some efficient strategies are omitted.

These examples of exclusions of alternative strategies from the base case are analogous to the examples of comparator omission identified in **Table 2**. Relevant strategies or comparators are omitted from the base-case analysis, meaning that what strategy is identified as optimal may be contingent on what analysts chose to compare.

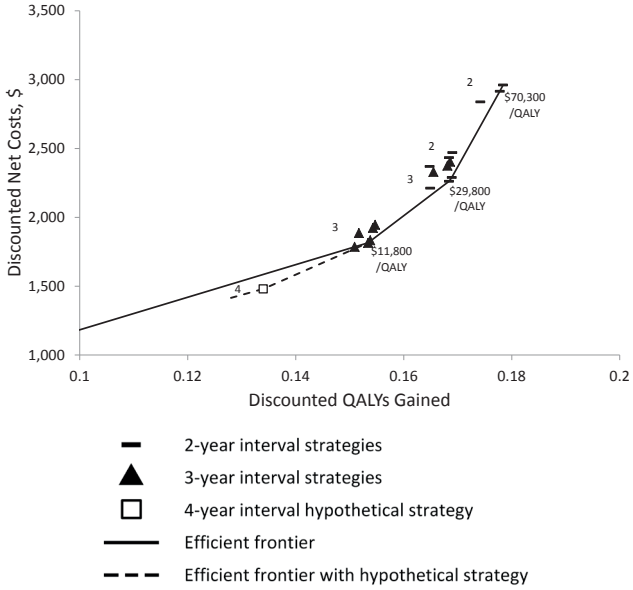


Figure 4. Cost-effectiveness plane from Mandelblatt et al. showing efficient frontier with dashes and triangles marking screening intervals of 2 and 3 years respectively and the inclusion of a hypothetical strategy with an interval of 4 years marked with a square.
QALY = quality-adjusted life year.

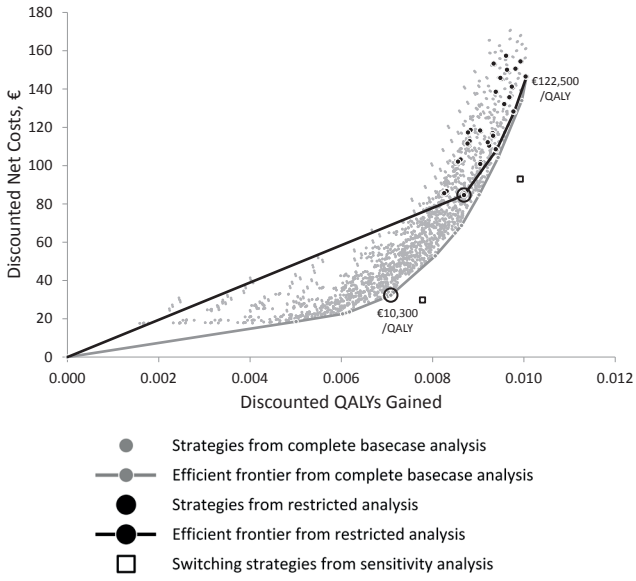


Figure 5. Cost-effectiveness plane from van Rosmalen et al. for all strategies in the original analysis shown in grey and for a restricted set of strategies with shorter screening intervals and longer screening ages shown in black, with the optimal strategies for a given of threshold of €20,000/QALY circled in each case and two switching strategies considered in a sensitivity analysis marked by the hollow squares.
QALY = quality-adjusted life year.

DISCUSSION

To date, there has been no systematic consideration of how cervical screening CEAs have addressed the expanded range of screening possibilities offered by HPV testing. This review adds to the literature by appraising the choice of comparator strategies in HPV screening CEAs with respect to the adequacy of ICER estimates. It has shown that many of these CEAs probably include insufficient comparator strategies to reliably estimate ICERs along the CEA-relevant portion of the efficient frontier. In particular, it is likely that many of the studies reviewed have not included strategies with sufficiently long screening intervals.

The omission of relevant comparators can lead to large errors in the ICER. The example from Goldie et al.[43] shows that a fourfold difference in the ICER can result from the omission of a longer interval. Such differences are not trivial and could lead to the mistaken adoption of cost-ineffective strategies. Similarly, the example of van Rosmalen et al.[8] shows that a restriction of the choice set could lead to the misidentification of an inefficient strategy as optimal and a 2.5-fold increase in screening costs.

Another problem identified in this review is that some CEAs exclude strategies from the base case, despite being shown to be relevant in secondary analyses. This could justifiably be considered a presentational issue rather than an analytical error. It should be acknowledged, however, that such reporting creates scope for misinterpretation and may confuse decision makers regarding which strategies are cost-effective and which are not.

That many of the studies reviewed appear to include insufficient comparators is clearly a critical finding. It is however important to acknowledge a number of important caveats. The first is that subjective judgment is involved when deciding whether sufficient comparators have been simulated or not. Although the methods and evidence supporting this review's conclusions have been documented, much relies on inference from previous studies regarding how the shape of the efficient frontier varies with screening intensity.

A second important caveat is that the conclusion of comparator omission depends in part on the cost-effectiveness threshold. The findings of this review would differ in particular cases if the threshold was different. For example, in the case of Goldie et al.[43] if a threshold of $100,000/QALY was used instead of the $50,000/QALY, then the $95,300/QALY strategy would be optimal. Because this strategy is supported by sufficient comparators, this would reverse the conclusion that Goldie et al.[43] probably includes insufficient comparators. Because very few countries have explicit thresholds, it is often unclear what the CEA-relevant portion of the efficient frontier is. In turn, this also means that the conclusion of comparator omission is less certain. An alternative perspective on this issue, however, is that if it is unclear what threshold applies, then it is arguably more incumbent on modelers to simulate a broad range of comparators because they should estimate the efficient frontier throughout the range of ICER values that might include the threshold.

A third important caveat concerns how policy relevance is defined and how this determines modeling choices. This analysis explicitly considers policy relevance from the perspective of CEA and assumes that all technically feasible strategies are candidates for

simulation. What decision makers consider policy relevant, however, may of course depend on other factors and not all strategies may be judged feasible. Certain strategies may be considered infeasible if they represent a large reduction in screening intensity relative to current guidelines or what screened populations are accustomed to, especially if the expected health gains are less than those of current practice.

Indeed, differences in the screening status quo may help explain the omission of longer screening intervals in many studies. No North American study considered intervals longer than 5 years, whereas 8 of 17 European CEAs did. US and Canadian screening guidelines recommend cytology every 3 years,[59,60] whereas the screening programs in the United Kingdom and The Netherlands feature 5-years intervals.[9] It seems plausible that analysts are reluctant to simulate strategies with intervals much longer than current services.

Although alternative policy perspectives are certainly valid, it is less clear to what extent they should determine what comparators can and cannot be assessed. Indeed, if there are constraints on what policies are deemed feasible, then it seems important that CEA should be free to assess the implied incremental cost of these restrictions. For instance, even if 5 years is the longest interval a decision maker judges feasible, it is probably still necessary to include a 6-year interval comparator against which to estimate its ICER. Without a 6-yearly comparator, the ICER of 5-yearly screening will likely be determined relative to no screening. Although this will probably yield a more favorable ICER, this seems unsatisfactory because no screening appears an even less feasible policy option than a 6-yearly interval. Allowing comparators to be excluded according to what is considered feasible means that ICERs may be arbitrarily determined by policymaker's preferences rather than incremental differences in costs and effects.

A limitation of this review is that it has primarily addressed the omission of comparators with longer screening intervals that can often be readily identified from kinked frontiers. The omission of other relevant comparators may not always be obvious. For example, the results of van Rosmalen et al.[8] indicate that shorter screening age ranges are also an important determinant of relevant screening intensity when specifying comparator strategies. This insight, however, could not be discerned from a CEA that does not include shorter screening age ranges. The same applies to the finding that strategies in which the order of cytology and HPV testing is switched around age 30 years perform better than do strategies without switching. These other aspects of comparator choice that have not been addressed in such detail here are also relevant.

The concerns raised here about the validity of ICERs based on limited comparisons are not novel.[4,61] As has been noted in the literature previously, however, the concern is not with the adequacy of CEA methods in principle, but their correct application in practice.[62] The role of comparator strategies is particularly important in the case of screening, because unlike in the case of drug interventions in which the number of comparator interventions is typically finite, the variety of possible screening strategies means the choice of comparators is at the modeler's discretion. It is notable that few of the analyses gave a justification for the range of strategies compared, possibly indicating that analysts are not explicitly aware of its sig-

nificance in determining estimates. The issues raised here regarding comparator omission apply equally to CEAs of other interventions with multiple possible comparators, including breast and colorectal screening, within which kinked frontiers can also be found.[63–65]

The observations of this review can be distilled into a simple three-item checklist to avoid some of the problems described above. First, check whether the efficient frontier features marked kinks rather than being gently curved because this may indicate the omission of relevant comparators. Second, ensure that the analysis includes sufficient comparator strategies against which to reliably estimate the ICERs of the CEA-relevant strategies; ideally, the screening interval should be progressively increased in annual increments so a range of ICERs is achieved that extends from well below the cost-effectiveness threshold to well above it. Finally, verify that all the relevant strategies simulated are included in the base-case analysis.

CONCLUSIONS

The importance of including relevant comparators has long been recognized, both in CEA in general and in the particular context of screening. Nevertheless, this review found that many CEAs of HPV screening would probably benefit from modeling more screening strategies, especially those of longer intervals. Hopefully, by drawing attention to specific examples within the literature, this review will refresh the attention of CEA analysts to the need to choose comparators carefully. Similarly, this analysis will hopefully help decision makers critically interpret CEA results an d to be aware of the implications of the choice of comparator strategies for ICER estimates. Doing so will bring easy-to-achieve enhancements to the reliability of cost-effectiveness evidence, which, in turn, will support better, more efficient screening policies for cervical cancer and other diseases.

REFERENCES

1. Weinstein MC. Principles of cost-effective resource allocation in healthcare organizations. Int J Technol Assess Health Care 1990;6:93–103.
2. Siegel JE, Weinstein MC, Torrance GW. Reporting cost-effectiveness studies and results. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, eds., Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.
3. Weinstein M, Zeckhauser R. Critical ratios and efficient allocation. J Public Econ 1973;2:147–57.
4. Torrance GW, Siegel JE, Luce BR. Framing and designing the cost effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, eds., Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.
5. L ăără E, Day N, Hakama M. Trends in mortality from cervical cancer in the Nordic countries: association with organised screening programmes. Lancet 1987;329:1247–9.
6. Kim JJ, Brisson M, Edmunds WJ, et al. Modeling cervical cancer prevention in developed countries. Vaccine 2008;26:K76–86.
7. Elfström KM, Arnheim-Dahlström L, von Karsa L, et al. Cervical cancer screening in Europe: quality assurance and organisation of programmes. Eur J Cancer 2015;51:950–68.
8. van Rosmalen J, de Kok IM, van Ballegooijen M. Cost-effectiveness of cervical cancer screening: cytology versus human papillomavirus DNA testing. Br J Obstet Gynaecol 2012;119:699–709.
9. Anttila A, von Karsa L, Aasmaa A, et al. Cervical cancer screening policies and coverage in Europe. Eur J Cancer 2009;45:2649– 58.
10. Adriaensen WJ, Matheï C, Buntinx FJ, et al. A framework provided an outline toward the proper evaluation of potential screening strategies. J Clin Epidemiol 2013;66:639–47.
11. Arbyn M, Ronco G, Anttila A, et al. Evidence regarding human papillomavirus testing in secondary prevention of cervical cancer. Vaccine 2012;30(S5):F88–99.
12. Cuzick J, Clavel C, Petry KU, et al. Overview of the European and North American studies on HPV testing in primary cervical cancer screening. Int J Cancer 2006;119:1095– 101.
13. Dillner J, Rebolj M, Birembaut P, et al. Long term predictive values of cytology and human papillomavirus testing in cervical cancer screening: joint European cohort study. BMJ 2008:337.
14. Diaz M, de Sanjose S, Ortendahl J, et al. Cost-effectiveness of human papillomavirus vaccination and screening in Spain. Eur J Cancer 2010;46:2973–85.
15. Coupé VM, Bogaards JA, Meijer CJ, et al. Impact of vaccine protection against multiple HPV types on the cost-effectiveness of cervical screening. Vaccine 2012;30:1813– 22.
16. International Monetary Fund. World Economic Outlook. Washington, DC: International Monetary Fund, 2013: Available from: http://www.imf.org/external/pubs/ft/weo/2013/01/pdf/text.pdf. [Accessed September 1, 2013].
17. Kulasingam SL, Kim JJ, Lawrence WF, et al. Cost-effectiveness analysis based on the atypical squamous cells of undetermined significance/ low-grade squamous intraepithelial lesion Triage Study (ALTS). J Natl Cancer Inst 2006;98:92–100.
18. Östensson E, Fröberg M, Hjerpe A, et al. Economic analysis of human papillomavirus triage, repeat cytology, and immediate colposcopy in management of women with minor cytological abnormalities in Sweden. Acta Obstet Gynecol Scand 2010;89:1316–25.
19. Lytwyn A, Sellors JW, Mahony JB, et al. Adjunctive human papillomavirus testing in the 2-year follow-up of women with low grade cervical cytologic abnormalities: a randomized trial and economic evaluation. Arch Pathol Lab Med 2003;127:1169–75.

20. Cochand-Priollet B, Cartier I, de Cremoux P, et al. Cost-effectiveness of liquid-based cytology with or without hybrid-capture II HPV test compared with conventional Pap smears: a study by the French society of clinical cytology. Diagn Cytopathol 2005;33:338–43.
21. Carozzi F, Cecchini S, Confortini M, et al. Role of P16 (INK4a) expression in identifying CIN2 or more severe lesions among HPV-positive patients referred for colposcopy after abnormal cytology. Cancer Cytopathol 2006;108:119–23.
22. Holmes J, Hemmett L, Garfield S. The cost-effectiveness of human papillomavirus screening for cervical cancer. Eur J Health Econ 2005;6:30–7.
23. Mühlberger N, Sroczynski G, Esteban E, et al. Cost-effectiveness of primarily human papillomavirus–based cervical cancer screening in settings with currently established Pap screening: a systematic review commissioned by the German Federal Ministry of Health. Int J Technol Assess Health Care 2008;24:184–92.
24. O'Mahony JF, van Rosmalen J, Zauber AG, et al. Multicohort models in cost-effectiveness analysis: why aggregating estimates over multiple cohorts can hide useful information. Med Decis Making 2013;33:407–14.
25. Campos NG, Castle PE, Schiffman M, et al. Policy implications of adjusting randomized trial data for economic evaluations: a demonstration from the ASCUS-LSIL triage study. Med Decis Making 2012;32:400–27.
26. Sroczynski G, Schnell-Inderst P, Mühlberger N, et al. Cost-effectiveness of primary HPV screening for cervical cancer in Germany—a decision analysis. Eur J Cancer 2011;47:1633–46.
27. Sroczynski G, Schnell-Inderst P, Mühlberger N, et al. Decision-analytic modeling to evaluate the long-term effectiveness and cost effectiveness of HPV-DNA testing in primary cervical cancer screening in Germany. GMS Health Technol Assess 2010;6:1–13.
28. van Ballegooijen M, van den Akker-van Marle M, Warmerdam P, et al. Present evidence on the value of HPV testing for cervical cancer screening: a model-based exploration of the (cost-) effectiveness. Br J Cancer 1997;76:651–7.
29. Kim JJ, Ortendahl J, Goldie SJ. Cost-effectiveness of human papillomavirus vaccination and cervical cancer screening in women older than 30 years in the United States. Ann Intern Med 2009;151:538–45.
30. Mandelblatt JS, Lawrence WF, Womack SM, et al. Benefits and costs of using HPV testing to screen for cervical cancer. JAMA 2002;287:2372–81.
31. Accetta G, Biggeri A, Carreras G, et al. Is human papillomavirus screening preferable to current policies in vaccinated and unvaccinated women? A cost-effectiveness analysis. J Med Screen 2010;17:181–9.
32. Balasubramanian A, Kulasingam SL, Baer A, et al. Accuracy and cost effectiveness of cervical cancer screening by high-risk human papillomavirus DNA testing of self-collected vaginal samples. J Low Genit Tract Dis 2010;14:185–95.
33. Berkhof J, Coupé VM, Bogaards JA, et al. The health and economic effects of HPV DNA screening in The Netherlands. Int J Cancer 2010;127:2147–58.
34. Berkhof J, de Bruijne MC, Zielinski GD, et al. Evaluation of cervical screening strategies with adjunct high-risk human papillomavirus testing for women with borderline or mild dyskaryosis. Int J Cancer 2006;118:1759–68.
35. Bidus MA, Maxwell GL, Kulasingam S, et al. Cost-effectiveness analysis of liquid-based cytology and human papillomavirus testing in cervical cancer screening. Obstet Gynecol 2006;107:997–1005.
36. Bistoletti P, Sennfält K, Dillner J. Cost-effectiveness of primary cytology and HPV DNA cervical screening. Int J Cancer 2008;122:372–6.
37. Burger EA, Ortendahl JD, Sy S, et al. Cost-effectiveness of cervical cancer screening with primary human papillomavirus testing in Norway. Br J Cancer 2012;106:1571–8.

38. Chen MK, Hung HF, Duffy S, et al. Cost-effectiveness analysis for Pap smear screening and human papillomavirus DNA testing and vaccination. J Eval Clin Pract 2011;17:1050–8.
39. Chow IH, Tang CH, You SL, et al. Cost-effectiveness analysis of human papillomavirus DNA testing and Pap smear for cervical cancer screening in a publicly financed health-care system. Br J Cancer 2010;103:1773–82.
40. Chuck A. Cost-effectiveness of 21 alternative cervical cancer screening strategies. Value Health 2010;13:169–79.
41. Coupé VM, de Melker HE, Snijders PJ, et al. How to screen for cervical cancer after HPV16/18 vaccination in The Netherlands. Vaccine 2009;27:5111–9.
42. Goldhaber-Fiebert JD, Stout NK, Salomon JA, et al. Cost-effectiveness of cervical cancer screening with human papillomavirus DNA testing and HPV-16,18 vaccination. J Natl Cancer Inst 2008;100:308–20.
43. Goldie SJ, Kim JJ, Wright TC. Cost-effectiveness of human papillomavirus DNA testing for cervical cancer screening in women aged 30 years or more. Obstet Gynecol 2004;103:619–31.
44. Kim JJ, Wright TC, Goldie SJ. Cost-effectiveness of alternative triage strategies for atypical squamous cells of undetermined signifi cance. JAMA 2002;287:2382–90.
45. Kim JJ, Wright TC, Goldie SJ. Cost-effectiveness of human papillomavirus DNA testing in the United Kingdom, The Netherlands, France, and Italy. J Natl Cancer Inst 2005;97:888–95.
46. de Kok IM, van Rosmalen J, Dillner J, et al. Primary screening for human papillomavirus compared with cytology screening for cervical cancer in European settings: cost effectiveness analysis based on a Dutch microsimulation model. BMJ 2012;344:e670.
47. Kulasingam SL, Rajan R, St Pierre Y, et al. Human papillomavirus testing with Pap triage for cervical cancer prevention in Canada: a cost effectiveness analysis. BMC Med 2009;7:69.
48. Legood R, Gray A, Wolstenholme J, et al. Lifetime effects, costs, and cost effectiveness of testing for human papillomavirus to manage low grade cytological abnormalities: results of the NHS pilot studies. BMJ 2006;332:79–85.
49. Maxwell GL, Carlson JW, Ochoa M, et al. Costs and effectiveness of alternative strategies for cervical cancer screening in military beneficiaries. Obstet Gynecol 2002;100:740–8.
50. Mittendorf T, Petry KU, Iftner T, et al. Economic evaluation of human papillomavirus screening in Germany. Eur J Health Econ 2003;4:209– 15.
51. Östensson E, Hellström AC, Hellman K, et al. Projected cost effectiveness of repeat high-risk human papillomavirus testing using self-collected vaginal samples in the Swedish cervical cancer screening program. Acta Obstet Gynecol Scand 2013;92:830–40.
52. Rogoza RM, Ferko N, Bentley J, et al. Optimization of primary and secondary cervical cancer prevention strategies in an era of cervical cancer vaccination: a multi-regional health economic analysis. Vaccine 2008;26(S5):F46–5 8.
53. Sherlaw-Johnson C, Philips Z. An evaluation of liquid-based cytology and human papillomavirus testing within the UK cervical cancer screening programme. Br J Cancer 2004;91:84–9 1.
54. Vijayaraghavan A, Efrusy MB, Goodman KA, et al. Cost-effectiveness of using human papillomavirus 16/18 genotype triage in cervical cancer screening. Gynecol Oncol 2010;119:237–42.
55. Vijayaraghavan A, Efrusy MB, Mayrand MH, et al. Cost-effectiveness of high-risk human papillomavirus testing for cervical cancer screening in Quebec, Canada. Can J Public Health 2010;101:220–5.
56. van den Akker-van ME, van Ballegooijen M, van Oortmarssen GJ, et al. Cost-effectiveness of cervical cancer screening: comparison of screening policies. J Natl Cancer Inst 2002;94:193–2 0 4.
57. Eurostat. Annual National Accounts 2015. Available from: http://ec.europa.eu/eurostat/web/national-accounts/data/main-tables.[Accessed July 21, 2015].

58. World Health Organization. WHO Commission on Macroeconomics and Health. Investing in Health for Economic Development: Report of the Commission on Macroeconomics and Health. Geneva: World Health Organization, 2001.
59. Moyer VA. Screening for cervical cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med 2012;156:880–91.
60. Canadian Task Force on Preventive Health Care. Recommendations on screening for cervical cancer. Can Med Assoc J 2013;185:35–45.
61. Palmer S, Raftery J. Opportunity cost. BMJ 1999;318:1551–2.
62. Maynard A, McDaid D. Evaluating health interventions: exploiting the potential. Health Policy 2003;63:215–26.
63. de Bock G, Vermeulen K, Jansen L, et al. Which screening strategy should be offered to women with BRCA1 or BRCA2 mutations? A simulation of comparative cost-effectiveness. Br J Cancer 2013;108:1579–86.
64. Woo PP, Kim JJ, Leung GM. What is the most cost-effective population based cancer screening program for Chinese women? J Clin Oncol 2007;25:617–24.
65. Vijan S, Hwang I, Inadomi J, et al. The cost-effectiveness of CT colonography in screening for colorectal neoplasia. Am J Gastroenterol 2007;102:380–90.

PART II

Reducing Harms

Chapter 3

The potential harms of primary human papillomavirus screening in over-screened women: a microsimulation study

Steffie K. Naber, Inge M.C.M. de Kok, Suzette M. Matthijsse, Marjolein van Ballegooijen

Cancer Causes Control. 2016 Mar; 27: 569-81.

ABSTRACT

Background: It is well acknowledged that HPV testing should not be performed at young age and at short intervals. Cytological screening practices have shown that over-screening, i.e. from a younger age and at shorter intervals than recommended, is hard to avoid. We quantified the consequences of a switch to primary HPV screening for over-screened women, taking into account its higher sensitivity but lower specificity than cytology.

Methods: The health effects of using the HPV test instead of cytology as the primary screening method were determined with the MISCAN-Cervix model. We varied the age women start screening and the interval between screens. In the sensitivity analyses, we varied the background risk for cervical cancer, the HPV prevalence, the discount rate, the triage strategy after cytology and the test characteristics of both cytology and the HPV test.

Results: For women screened 5-yearly from age 30, 32 extra deaths per 100,000 simulated women were prevented when switching from primary cytology to primary HPV testing. For annual screening from age 20, such a switch resulted in 6 extra deaths prevented. It was associated with 9,044 more positive primary screens in the former scenario versus 76,480 in the latter. Under all conditions, for women screened annually, switching to HPV screening resulted in a net loss of quality-adjusted life years.

Conclusion: For over-screened women, the harms associated with a lower test specificity outweigh the life years gained when switching from primary cytology to primary HPV testing. The extent of over-screening should be considered when deciding on inclusion of primary HPV screening in cervical cancer screening guidelines.

INTRODUCTION

In several Western countries, cytological screening has considerably reduced the cervical cancer incidence and mortality over the past four decades.[1] Nevertheless, even in countries with a nationwide screening program, women still die from cervical cancer. Although most deaths occur after age 30 and in women who did not adequately participate in screening, some deaths occur at young age and in women who recently received a negative test result (which suggests it was false negative).[2-4] Therefore, clinicians may tend to screen more frequently than recommended.[5]

Ever since infection with the human papillomavirus (HPV) was found to be a necessary condition for developing cervical cancer,[6,7] testing for the presence of high-risk HPV types (i.e., carcinogenic types) has received much attention. A summary of meta-analyses estimated that the HPV test has a 23% (95% CI: 13 to 33%) higher sensitivity, but a 6% (95% CI: 4 to 8%) lower specificity than cytology for detecting high-grade lesions and cervical cancer.[8] Cost-effectiveness analyses based on these findings have shown that in well-controlled screening situations primary HPV screening is likely to be more effective, as well as more cost-effective than primary cytology.[9,10] Therefore, many countries are considering a switch from primary cytology to primary HPV screening. In the US, co-testing (i.e. cytology combined with HPV testing) is already recommended, and Australia and the Netherlands are preparing a switch from primary cytology to primary HPV screening.[11-13]

For primary cytology it is known that over-screening, here defined as screening from a younger age or at shorter intervals than recommended, is neither required to detect progressive lesions in an early phase nor desired as it detects many regressive lesions. Unavoidably, it also involves more false-positive test results, adding to the psychological stress women may experience from having a positive test and being referred for colposcopy.[14] In addition, the costs of over-screening are substantial; amounting to approximately 0.5-1 billion USD per year for the US healthcare system, while yielding little or no health gains.[15]

Because of its lower specificity to detect clinically relevant lesions, avoiding over-screening is even more essential for HPV screening than for cytology screening. The vast majority of HPV infections clear spontaneously, especially at young age.[16] Detecting these infections leads to unnecessary triage situations or referrals to colposcopy. For over-screened women, switching to HPV (co-)testing may therefore do more harm than good.

Guidelines driven by rational decision making tend to restrict cytology screening–and HPV screening even more so. The US guidelines currently recommend cervical screening in women aged 21 to 65 years with an interval of 3 or 5 years (dependent on both age and test).[17] In European guidelines, primary HPV screening is recommended for women aged ≥35 and discouraged for those below the age of 30.[18] In the Netherlands, primary HPV screening will be offered from age 30 to 65 every 5-10 years, and in Australia from age 25 to 69 every 5 years.[11,13] Unfortunately, also for HPV screening, having well-considered screening policy recommendations will not guarantee that women are screened accordingly.

A recent US study showed that over 68% of physicians would recommend another cytological test in 1 or 2 years where the guidelines recommend a 3-year interval.[19] After a negative co-test, 67-94% of clinicians recommended a shorter screening interval than suggested by US guidelines.[20] Several European countries also have reported considerable over-screening.[21] In summary, large proportions of women are being over-screened with cytology, and this is likely to continue when HPV screening is implemented.

Notwithstanding these facts, HPV testing is, for good reasons, increasingly often included in primary screening recommendations. However, despite its lower specificity, we are unaware of intensified efforts to minimize the level of over-screening. In this study, we aim to quantify the harms and benefits of introducing primary HPV screening for women with diverse screening behaviors, with age of first screen ranging from 20 to 30 years, and screening interval from 1 to 5 years. These scenarios cover both recommended schedules and observed levels of over-screening. The results of this study show the effects of introducing HPV screening for over-screened women, as well as for those who adhere to guidelines. Although the model was based on Dutch data, the resulting outcomes are important for all over-screened women, regardless of where they live. Since it seems too early to draw conclusions on the effect of switching to HPV screening in over-screened women who have been vaccinated, this analysis only considers unvaccinated cohorts.

METHODS

Health effects of different screening scenarios were estimated using the MISCAN-Cervix model, which is described in more detail in the **Model appendix**.[22]

MISCAN-Cervix model

MISCAN-Cervix is a microsimulation model in which a large study population with individual life histories is generated. In all of the analyses presented here, we simulated a 20-year-old cohort of 100 million women with life expectancy as observed in the Netherlands,[23] which was not affected by HPV vaccination (neither directly nor through herd immunity). A fraction of these women will acquire HPV infections and/or develop cervical intraepithelial neoplasia (CIN) lesions. If these precursors progress to cervical cancer, the result may be death. Screening can detect the disease, which can then be treated at an earlier stage. As a result, cervical cancer death may be prevented or postponed.

In the model, the disease development is in seven sequential stages: high-risk HPV infection, three pre-invasive stages (CIN 1, 2, and 3), and three invasive stages (International Federation of Gynecology and Obstetrics (FIGO) stages IA, IB and II or worse). Whereas pre-invasive and FIGO IA stages can only be diagnosed by screening, because at these stages women are assumed to be symptom-free, FIGO IB or worse can also be clinically diagnosed. As precursors are usually not progressive,[24] over 90% of modeled HPV infections clear without ever resulting in neoplasia and most pre-invasive lesions regress spontaneously. In

the hypothetical situation without competing other-cause mortality, undetected preclinical invasive neoplasia will always progress to clinical cancer. CIN grades 1 and 2 can develop in the absence of a high-risk HPV infection; in that case the lesion will always regress. CIN grade 3 or worse can only develop if a high-risk HPV infection is present.

Triage strategies

For primary HPV screening and primary cytology we used a cost-effective triage strategy, as published previously.[9] Primary cytological test results classified as atypical squamous cells of undetermined significance (ASCUS) or low-grade squamous intraepithelial lesion are immediately followed by an HPV test using the same material. A positive primary HPV test is immediately followed by cytology using the same material. If no cytological abnormalities are found, another cytological test is performed after 6 months.

Although the latter strategy will also be implemented in the Dutch screening program in 2017, triage strategies were not selected based on guidelines or current practice. Instead, we decided to select strategies based on cost-effectiveness, such that inefficiencies in triage strategies would not dilute or exaggerate the effect of switching to HPV screening. The triage practices of over-screened women are unknown and might be very heterogeneous. It seems unlikely that women who do not follow primary screening guidelines, do follow the exact triage recommendations. We therefore chose to simulate a relatively simple triage strategy for both primary tests, and to focus on the number of positive primary tests (i.e., those that require follow-up) instead of on the number of triage tests.

Screening scenarios

We simulated 12 cohorts with different screening behaviors, varying the age at which women start screening (20, 25, or 30 years) and the frequency with which they get tested (every 1, 2, 3, or 5 years). In all scenarios, screening was assumed to end at or before the age of 65.[17,25] The resulting outcomes are only relevant for women having the screening behavior as modeled, and should not be translated to an entire population.

Assumptions for screening and treatment

Table 1 presents the base-case assumptions for screening. We assumed the sensitivity of cytology (that is, the probability that the result is at least ASCUS) to be 40% for true stage CIN grade 1, 50% for CIN grade 2 and 75% for CIN grade 3 or cancer.[26] In the model calibration, the sensitivity of testing for at least high-grade squamous intraepithelial lesion (HSIL), the cytological cut-off for referral to colposcopy, and therefore for detection, was estimated to be 4% for CIN grade 1, 18% for CIN grade 2, 56% for CIN grade 3 and 60% for cervical cancer. Furthermore, the specificity of cytology was estimated to be 97.6% based on Dutch data.[27] Based on the observed difference in CIN grade 3 or cancer detection rates between cytology and the HPV test, we assumed the sensitivity of the HPV test to be 94% for a high-risk HPV infection.[28] As we assumed that cervical cancer can only develop if an HPV infection is present, the sensitivity for cervical cancer is also 94%. The overall sensitivity

for CIN lesions is lower and depends on the age-specific prevalence of HPV infections in CIN lesions. In the model, the specificity for detecting high-risk HPV infections was assumed to be 100%. A probable (but unknown) lack of specificity was accounted for by the inclusion of fast-clearing infections, in concordance with HPV clearing studies.[29,30]

Table 1. Base-case model inputs and variations in the sensitivity analyses.

Parameter	Base-case value	Alternative value(s)
Background risk of cervical cancer mortality	5 per 100,000 life years	10 per 100,000 life years
HPV prevalence in women without CIN grade 2 or worse[a]	Low	High[b]
Sensitivity of cytology		
Probability of at least ASCUS (at least triage) for:		
CIN grade 1	40%[26]	32%
CIN grade 2	50%[26]	40%
CIN grade 3 or worse	75%[26]	60%
Probability of at least HSIL (referral for colposcopy) for:		
CIN grade 1	4%[c]	3%
CIN grade 2	18%[c]	14%
CIN grade 3	56%[c]	45%
Cervical cancer	60%[c]	48%
Specificity of cytology[d]	97.6%[c]	95.2%
Sensitivity of HPV test[e]	94%[28]	85%[8], 100%[8]
Specificity of HPV test	100%[f]	Not varied as such[g]
Discounting	3%[31]	0%, 5%

HPV = human papillomavirus; CIN = cervical intraepithelial neoplasia; ASCUS = atypical squamous cells of undetermined significance; HSIL = high grade squamous intraepithelial lesion.

[a] Depends on age, age-dependency was not varied.

[b] The number of false-positive referrals to colposcopy and CIN 1 lesions was doubled.

[c] Value was determined in model calibration.

[d] Probability of a normal test result in women without CIN or cancer.

[e] Probability to detect an HPV infection, regardless of whether a CIN lesion or cancer is present.

[f] A possible lack of specificity was modeled by including fast-clearing HPV infections.

[g] As a lower specificity of the HPV test corresponds with a higher prevalence of harmless HPV infections in the model, this parameter was not varied.

Detection and management of pre-invasive lesions, including treatment if necessary, was assumed to lead to a 100% cure rate. However, new HPV infections and recurring CIN lesions after CIN treatment cannot be excluded. For invasive cancer, we determined age-specific and stage-specific survival probabilities based on data from the Netherlands Cancer Registry.[31] Since cancers detected by screening are usually at a less advanced stage than clinically diagnosed ones, women have a higher chance to survive them. If an invasive

cancer is screen-detected, the probability to die from cervical cancer is reduced by 89.4%, 50% and 20% when detected in FIGO stages IA, IB and II or worse, respectively.

Table 2 presents the utility losses assumed in the base-case scenario. A small (psychological) loss in quality of life is assumed for attending a screen (including waiting for the result) and for being in triage (including attending follow-up screenings). Larger losses in quality of life are assumed for being diagnosed and treated for CIN or cancer and for having a terminal stage of cervical cancer. We based the utility losses on nationally and internationally published data.[32-35]

Table 2. Model inputs regarding the utility loss due to screening, treatment and terminal care.

	Disutility	Duration	Quality-adjusted time lost
Screening[34]			
Primary screening	0.005	2 weeks	2 hours
Being in triage	0.005	0.5 year[a]	22 hours
False-positive referral	0.005	0.5 year	22 hours
Treatment of pre-invasive lesions[33]			
CIN grade 1	0.03	0.5 year	6 days
CIN grade 2 or 3	0.07	1 year	26 days
Cancer treatment[32,33] **and terminal care**[35]			
FIGO stage I	0.062	5 years	4 months
FIGO stage II+	0.280	5 years	17 months
Terminal care	0.740	1 year	9 months

CIN = cervical intraepithelial neoplasia; FIGO = International Federation of Gynecology and Obstetrics.

[a] Time between primary and triage test is 6 months.

Base-case analysis

For every scenario, we first estimated health effects of both primary cytology and primary HPV testing as compared to the situation without screening. Then, differences in health effects between these two interventions were explored. A first indication of the harm-benefit balance of introducing primary HPV testing is given by the number of additional positive primary screens (i.e., at least ASCUS for cytology screening and HPV-positive for HPV screening) that is required to prevent one additional cervical cancer death. As women with a positive primary screen require follow-up in terms of triage or colposcopy, we refer to this outcome measure as 'Number Needed to Follow-up', or NNF.

Comparing the life years lost to cervical cancer between the two interventions yields the number of life years gained by switching to the more sensitive primary HPV testing. Similarly, the difference in total disutility due to screening and treatment caused by these interventions can be computed. As the number of quality-adjusted life years (QALYs) gained combines these positive and negative effects of screening, this outcome measure was used

to compare the total health effects of primary HPV screening with those of primary cytology. Health effects were discounted to the year in which all women are 20 years old, using an annual rate of 3%.[36]

Sensitivity analyses

Some model parameters may have a non-negligible level of uncertainty, while others differ among countries or geographical regions. In one-way sensitivity analyses, we varied these types of parameters, covering for high-income countries, if they would influence the difference in health effects between primary HPV screening and primary cytology (**Table 1**).

Among Dutch women, the assumed background risk of dying from cervical cancer is relatively low (5 deaths per 100,000 life years). We have doubled this risk to determine the effects for countries with a higher risk.

To observe the effect of a higher prevalence of harmless HPV infections, we have doubled the number of referrals that did not result in the detection of a clinically relevant lesion (i.e., CIN 2+). Detecting more harmless HPV infections implicitly corresponds with a lower clinically relevant specificity of HPV testing.

Presumably, the high level of quality assurance in the Netherlands contributes to a relatively high quality of cytology compared to less controlled situations. To explore the impact of switching to HPV testing for settings with a lower quality of cytology, the sensitivity of cytology in both primary and triage testing was reduced by 20% in one of the sensitivity analyses. In another sensitivity analysis, the lack of specificity of cytology in both primary and triage testing was doubled from 2.4% to 4.8%.

Some uncertainty exists about the sensitivity of the HPV test, which may also vary between tests and situations. A summary of meta-analyses found that the relative sensitivity of the HPV test as compared to cytology is 1.23 (95% CI: 1.13 to 1.33). Based on this confidence interval, the sensitivity of the HPV test was assumed to be 85% in one of the sensitivity analyses, and 100% in another.[8] As these are assumed probabilities to detect an HPV infection, and women with a CIN lesion are not necessarily HPV infected, the sensitivity for CIN lesions is still lower than 100% in the latter scenario.

In another sensitivity analysis, the triage strategy after a positive cytological test was adjusted to reflect current Dutch screening guidelines. According to these guidelines, women with HSIL are directly referred for colposcopy and women with ASCUS or low-grade intraepithelial lesion (LSIL) are invited for cytology and HPV triage after 6 months. Women testing HSIL or ASCUS/LSIL and HPV-positive at this point in time will be referred for colposcopy, and women testing either ASCUS/LSIL or HPV-positive will be invited for another cytological test at 18 months.

Lastly, as reported discount rates vary from 0% to 5%, we also present the health effects when using an annual discount rate of 0% and of 5%.

RESULTS

Base-case analysis

For the 12 different screening scenarios considered, **Table 3** shows the impact of replacing primary cytology with primary HPV screening. The numbers are based on the undiscounted results of primary cytology and primary HPV screening compared to the situation without screening, as displayed in **Appendix Tables 1** and **2**, respectively. Although in practice it is very unlikely that the start age is well-controlled while the screening interval is not, we first discuss the effects of switching to HPV testing in women who start screening at age 30 and have repeated testing at intervals that are either recommended or shorter than recommended. Then, we discuss the effects of switching for women who are not only screened more frequent than recommended, but also from a younger age.

Frequent screening from age 30

For 5-yearly screening starting at age 30, replacing primary cytology with primary HPV screening reduced the number of cervical cancer deaths by 32 per 100,000 simulated women, which was a reduction of 27% (**Figure 1**, **Table 3**). This reduction was achieved at the expense of 9,044 more positive primary screens per 100,000 women (+34%), resulting in 2,572 more referrals to colposcopy (+29%). With annual screening in the same age range, switching to primary HPV screening would prevent only 7 extra deaths per 100,000 women (-9%), while positive primary screens would increase by 14,271 (+14%), and referrals to colposcopy by 3,477 (+19%). The (discounted) NNF was 769 in the first scenario versus 11,880 in the latter, more intensive one (**Table 4**).

Frequent screening from age 20

With annual screening starting at the age of 20 instead of 30, switching from primary cytology to primary HPV screening resulted in similar benefits (i.e., 6 additional deaths prevented per 100,000 women (-9%)). However, the number of women with a positive screen test increased by 76,480 instead of by 14,271 per 100,000 women. The NNF equaled 60,133, which was more than 5 times the NNF of switching in case of annual screening from age 30, and more than 78 times the NNF of switching in case of 5-yearly screening from age 30.

Changes in QALYs

Table 5 shows the QALYs gained (or lost) by switching from primary cytology to primary HPV screening for the diverse screening behaviors. Under base-case assumptions, a substantial number of QALYs were gained for women who were screened every 5 years from age 30. For more intensively screened women, the benefit of switching to HPV screening was uncertain. For women screened annually or biennially from any age, or triennially from age 20 or 25, replacing primary cytology with primary HPV testing even resulted in a net health loss.

Table 3. The impact of replacing primary cytology with primary HPV screening for 12 different screening scenarios.[a]

Screening interval	Start age	Primary screens[b]	Positive primary screens	Referrals to colposcopy	False-positive referrals (no CIN detected)	CIN 1	CIN 2	CIN 3	Cervical cancer cases	Cervical cancer deaths
	30	+1,014 (+0%)	+9,044 (+34%)	+2,572 (+29%)	+308 (+42%)	+1,651 (+59%)	+722 (+36%)	-71 (-2%)	-114 (-29%)	-32 (-27%)
5 years	25	+1,455 (+0%)	+17,741 (+54%)	+3,743 (+31%)	+497 (+50%)	+2,311 (+58%)	+1,040 (+36%)	-63 (-2%)	-116 (-32%)	-32 (-27%)
	20	+1,807 (+0%)	+22,293 (+59%)	+4,747 (+33%)	+581 (+51%)	+3,079 (+60%)	+1,282 (+37%)	-153 (-3%)	-117 (-33%)	-32 (-27%)
	30	+1,352 (+0%)	+11,375 (+30%)	+2,932 (+26%)	+442 (+40%)	+2,247 (+55%)	+652 (+25%)	-384 (-12%)	-67 (-20%)	-19 (-16%)
3 years	25	+2,136 (+0%)	+24,759 (+52%)	+4,332 (+28%)	+754 (+49%)	+3,185 (+54%)	+950 (+25%)	-530 (-12%)	-66 (-24%)	-18 (-18%)
	20	+2,909 (+0%)	+32,648 (+58%)	+5,698 (+30%)	+931 (+51%)	+4,403 (+56%)	+1,184 (+25%)	-792 (-17%)	-65 (-27%)	-17 (-19%)
	30	+1,980 (+0%)	+12,642 (+23%)	+3,220 (+24%)	+609 (+38%)	+2,756 (+49%)	+437 (+14%)	-566 (-18%)	-41 (-15%)	-12 (-13%)
2 years	25	+3,204 (+0%)	+31,715 (+47%)	+4,780 (+25%)	+1,072 (+47%)	+3,891 (+49%)	+631 (+14%)	-798 (-19%)	-39 (-19%)	-11 (-14%)
	20	+4,023 (+0%)	+45,199 (+59%)	+6,316 (+27%)	+1,373 (+51%)	+5,367 (+50%)	+734 (+13%)	-1,142 (-26%)	-39 (-18%)	-12 (-13%)
	30	+3,719 (+0%)	+14,271 (+14%)	+3,477 (+19%)	+1,093 (+35%)	+3,113 (+34%)	-140 (-4%)	-584 (-22%)	-18 (-8%)	-7 (-9%)
1 year	25	+6,113 (+0%)	+51,316 (+43%)	+5,361 (+21%)	+2,046 (+47%)	+4,330 (+34%)	-200 (-4%)	-811 (-23%)	-17 (-10%)	-7 (-9%)
	20	+8,211 (+0%)	+76,480 (+55%)	+7,259 (+22%)	+2,675 (+50%)	+6,096 (+35%)	-400 (-6%)	-1,108 (-35%)	-17 (-10%)	-6 (-9%)

CIN = cervical intraepithelial neoplasia.

[a] The table shows undiscounted numbers per 100,000 simulated women, with percentage changes between brackets.

[b] As compared to primary cytology, the number of primary screens is slightly higher for HPV screening (i.e., less than 1%) because it detects more (progressive) CIN lesions, resulting in fewer women being diagnosed with cervical cancer. Whereas women who have been diagnosed with a CIN lesion are assumed to be referred back to routine screening, those with cervical cancer are not.

Note: Numbers are differences between primary cytology and primary HPV screening, shown separately in **Appendix Tables 1** and **2**.

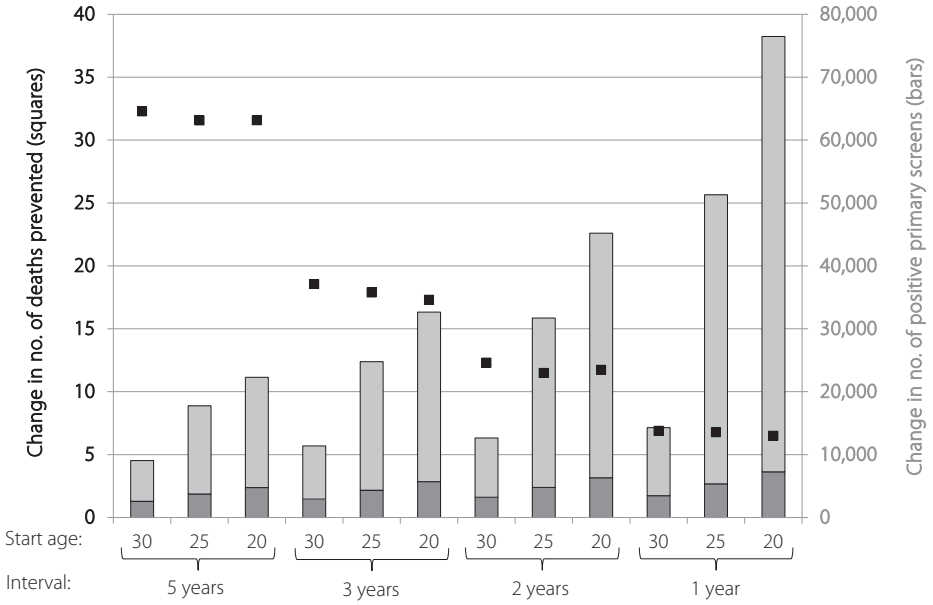


Figure 1. Simulated increase in lifetime number of deaths from cervical cancer prevented (left axis) and positive primary screens (right axis) when primary cytology is replaced with primary HPV screening. The increase in positive primary tests is split up in referrals to colposcopy (dark grey) and non-referrals to colposcopy (light grey). Undiscounted results for different start ages and intervals of screening are given per 100,000 women.

Sensitivity analyses

In all sensitivity analyses, primary HPV screening prevented more cervical cancer deaths than did primary cytology. In most scenarios, this occurred at the expense of more positive screens, and the NNF increased quite rapidly with the intensity of the screening scenario (**Table 4**). Only when the specificity of cytology was assumed to be lower (95.2% instead of 97.6%), for some levels of over-screening the number of positive screens decreased with the shift to primary HPV testing. The discount rate appeared to have the largest impact on the NNF.

In the sensitivity analyses, switching to primary HPV testing resulted in fewer QALYs gained in the case of more intensive screening. Overall, for a given level of over-screening, whether QALYs were gained or lost did not vary substantially among the sensitivity analyses. Generally, switching was favorable for women screened every five years, and unfavorable for those screened annually or biennially. However, when cytology was triaged as is currently recommended in the Dutch screening program or when health effects were not discounted, switching to HPV screening also resulted in QALYs gained for women screened biennially from age 30. For women screened every 5 years from age 20-25, QALYs were lost when the HPV prevalence was increased and when results were discounted at an annual rate of 5%.

Table 4. Number of additional positive primary screens per additionally prevented cervical cancer death (NNF) when primary cytology is replaced with primary HPV screening, for the base case and eight sensitivity analyses.[a]

Screening interval	Start age	Base case	Background risk of cervical cancer mortality ↑	HPV prevalence in CIN 1 or less[b] ↑	Sensitivity of cytology ↓	Specificity of cytology ↓	Sensitivity of HPV test ↑	Sensitivity of HPV test ↓	Cytology triaged as in Dutch program	No discounting	5% discounting
5 years	30	769	399	887	503	NA[c]	805	761	657	280	1,256
	25	1,589	811	1,788	993	502	1,628	1,638	1,360	562	2,692
	20	2,065	1,057	2,352	1,309	747	2,117	2,108	1,772	706	3,603
3 years	30	1,889	969	2,202	1,190	NA[c]	2,047	1,690	1,532	613	3,223
	25	4,443	2,289	4,997	2,712	1,009	4,725	4,213	3,645	1,385	7,927
	20	6,444	3,275	7,324	3,827	1,980	6,907	6,088	5,282	1,886	12,056
2 years	30	3,865	2,006	4,531	2,545	NA[c]	4,360	3,197	2,983	1,027	7,182
	25	10,526	5,341	11,779	6,443	1,346	11,521	9,381	8,140	2,760	20,582
	20	15,405	7,764	17,331	9,506	4,229	16,723	13,750	11,941	3,850	32,252
1 year	30	11,880	6,220	13,731	9,024	NA[c]	14,088	8,562	NA[d]	2,073	27,408
	25	36,576	18,503	39,954	28,544	NA[c]	40,654	30,861	NA[d]	7,570	86,763
	20	60,133	29,372	65,790	45,416	NA[c]	66,783	50,634	NA[d]	11,788	156,829

HPV = human papillomavirus; CIN = cervical intraepithelial neoplasia; NA = not applicable

[a] Numbers were discounted with an annual rate of 3%, unless stated otherwise.

[b] The number of women with a false-positive result or CIN 1 was doubled to account for a higher HPV prevalence among these women.

[c] The number of positive primary screens decreased with switching to HPV screening.

[d] The current Dutch screening program involves triage testing at 18 months after the primary test, which, for annual screening, interferes with the next screening round.

DISCUSSION

Even in countries with carefully constructed screening guidelines, women may be over-screened. As for over-screened women the risk for cervical cancer is already strongly reduced with primary cytology, the gains of switching to primary HPV screening are expected to be relatively small. Indeed, our analysis predicted that while switching would prevent 32 deaths per 100,000 women who are screened every 5 years, only 6-7 deaths would be averted in those screened annually. In the latter group, the increase in positive tests and subsequent follow-up procedures even resulted in a net loss in health.

Because the same conclusion was reached in all of the sensitivity analyses, it is likely generalizable to other developed countries. The lower the ratio of HPV prevalence to cervical cancer mortality risk, the less harmful HPV testing will be for over-screened women. The sensitivity analysis in which we doubled the lifetime risk of dying from cervical cancer

Table 5. Simulated change in number of QALYs gained and percentage change when primary cytology is replaced with primary HPV screening, for the base case and eight sensitivity analyses.[a]

Screening interval	Start age	Base case	Background risk of cervical cancer mortality ↑	HPV prevalence in CIN 1 or less[b] ↑	Sensitivity of cytology ↓	Specificity of cytology ↓	Sensitivity of HPV test ↑	Sensitivity of HPV test ↓	Cytology triaged as in Dutch program	No discounting	5% discounting
5 years	30	+116 (+4%)	+319 (+5%)	+101 (+3%)	+265 (+9%)	+115 (+4%)	+130 (+4%)	+80 (+2%)	+175 (+20%)	+667 (+5%)	+27 (+2%)
	25	+57 (+2%)	+258 (+4%)	+34 (+1%)	+203 (+7%)	+58 (+2%)	+63 (+2%)	+37 (+1%)	+124 (+15%)	+591 (+4%)	-26 (-2%)
	20	+16 (+1%)	+214 (+3%)	-19 (-1%)	+153 (+5%)	+16 (+1%)	+17 (+1%)	+3 (+0%)	+98 (+13%)	+559 (+4%)	-70 (-6%)
3 years	30	+4 (+0%)	+105 (+2%)	-16 (-0%)	+83 (+3%)	+4 (+0%)	+4 (+0%)	-2 (-0%)	+72 (+8%)	+254 (+2%)	-27 (-2%)
	25	-67 (-2%)	+24 (+0%)	-99 (-3%)	-6 (-0%)	-66 (-2%)	-75 (-2%)	-59 (-2%)	+16 (+2%)	+157 (+1%)	-89 (-7%)
	20	-125 (-4%)	-37 (-1%)	-175 (-6%)	-71 (-2%)	-124 (-4%)	-138 (-4%)	-111 (-4%)	-15 (-2%)	+88 (+1%)	-147 (-13%)
2 years	30	-52 (-2%)	+2 (+0%)	-76 (-2%)	-19 (-1%)	-52 (-2%)	-57 (-2%)	-49 (-1%)	+34 (+4%)	+56 (+0%)	-55 (-4%)
	25	-133 (-4%)	-86 (-1%)	-172 (-5%)	-114 (-4%)	-132 (-4%)	-144 (-4%)	-117 (-4%)	-20 (-2%)	-55 (-0%)	-125 (-10%)
	20	-196 (-6%)	-148 (-2%)	-257 (-9%)	-189 (-6%)	-195 (-6%)	-210 (-7%)	-174 (-6%)	-41 (-7%)	-112 (-1%)	-191 (-19%)
1 year	30	-127 (-4%)	-106 (-2%)	-155 (-5%)	-128 (-4%)	-125 (-4%)	-135 (-4%)	-115 (-4%)	NA[c] (-)	-154 (-1%)	-99 (-8%)
	25	-238 (-8%)	-218 (-3%)	-283 (-10%)	-254 (-8%)	-237 (-8%)	-253 (-8%)	-214 (-7%)	NA[c] (-)	-290 (-2%)	-197 (-19%)
	20	-334 (-12%)	-315 (-5%)	-407 (-15%)	-360 (-13%)	-332 (-12%)	-354 (-13%)	-303 (-11%)	NA[c] (-)	-390 (-3%)	-293 (-36%)

HPV = human papillomavirus; CIN = cervical intraepithelial neoplasia; NA = not applicable.

[a] The table shows numbers per 100,000 simulated women, with percentage changes between brackets. Numbers were discounted with an annual rate of 3%, unless stated otherwise.

[b] The number of women with a false-positive result or CIN 1 was doubled to account for a higher HPV prevalence among these women.

[c] The current Dutch screening program involves triage testing at 18 months after the primary test, which, for annual screening, interferes with the next screening round.

showed that it would still be harmful if this ratio would be twice as low as in the Netherlands though. In countries with an even lower HPV prevalence to cervical cancer mortality risk ratio, switching to HPV testing might be beneficial for over-screened women. In the US, however, both HPV prevalence and cervical cancer mortality are comparable to the Netherlands.[37,38] In most European countries, cervical cancer mortality is higher,[39] but HPV prevalence is also (up to) twice as high.[38]

Obviously, the goal of a cancer screening program is to decrease the disease's incidence and mortality rate. Because in every simulated scenario switching from primary cytology to primary HPV screening reduced the number of cervical cancer cases and deaths, one could argue that primary HPV screening should always be preferred. This would indeed be true if being in triage, being referred to colposcopy, and being treated for CIN would not be associated with losses in quality of life. However, the health-related burden of these events is a drawback of screening that should not be overlooked.[40,41]

A number of randomized controlled trials (RCTs) have compared primary cytology screening to either HPV screening alone or to HPV screening combined with cytology.[42-45] In these RCTs, HPV screening resulted in a higher detection rate of CIN lesions and an improved protection against cervical cancer.[46] CEAs based on these findings showed that primary HPV screening with an interval of at least three years is cost-effective for women above age 30.[9,47] We showed that the effectiveness is questionable if this cannot be guaranteed. In this regard, data from a US population-based registry showed that recommending 3-yearly cytology screening resulted in a median time between two consecutive smears of 1.87 years in 2011.[48] There is no reason to assume that guidelines regarding primary HPV screening would be followed more closely. In fact, a study from 2010 found a lower adherence to guidelines after a negative co-test as compared to after a negative cytological test.[19] Although co-testing is intended for women who want to extend their screening interval from 3 to 5 years, many clinicians provide it on an annual basis.[19]

Switching to HPV screening could be considered more effective for women with that level of over-screening for which HPV screening was associated with a net health benefit, but this would not necessarily be more cost-effective. However, the decision to include primary HPV screening in national screening guidelines should take into account its population-level cost-effectiveness. If only a relatively small number of women are over-screened, then switching to HPV screening may well be (very) cost-effective on a population level. In the Netherlands, given the small number of smears taken outside the screening program,[49] it is expected to be cost-effective.

Strengths and limitations

Even though earlier research showed that primary HPV screening is more cost-effective than primary cytology for women who adhere to screening guidelines,[10,50] this is the first study to quantify its harms and benefits for over-screened women. As over-screening practices are likely to remain, these results are relevant to any country considering recommending primary HPV screening, either alone or as a co-test.

Our study also has some limitations. First of all, our model is based on Dutch data. Although it might have been better to adjust the model for every single country, we did vary those country-specific parameters that would influence the conclusion. For example, we increased the HPV prevalence level to estimate effects for high HPV prevalence countries such as Denmark.[51] We did not modify the prevalence age distribution as the peak between the ages of 20 and 30 has also been observed in other European countries and in the United States.[51,52]

Although we varied test characteristics to explore the effect of switching to HPV screening for different settings, the ranges considered are not representative for low- and middle-income countries, where sustaining cytology programs of sufficient quality is often difficult.[53,54] As the test characteristics are only one of many factors that may be different in those countries, separate analyses are needed for these situations.

Meta-analyses have shown that removal of CIN lesions carries an increased risk of having pre-term births.[55,56] We did not include this potential harm because estimates of the impact on a woman's quality of life are unavailable. If we would have accounted for this in our analyses, in over-screened women even more QALYs would have been lost by switching to primary HPV screening.

Although there are numerous possible triage strategies for cytology and HPV testing, in the base-case analysis we only considered two that were found to be cost-effective in a previous analysis.[9] In a sensitivity analysis we did explore the impact of switching from the less efficient cytology screening strategy that is currently recommended in the Netherlands to the cost-effective HPV screening strategy that will be implemented in 2017. When these less efficient cytology practices were assumed, switching to HPV testing was obviously more beneficial. Nevertheless, it still resulted in a net health loss for women screened biennially from age 20-25 or triennially from age 20 (effects for annually screened women were not evaluated for this triage strategy). If future triage practices would be much more efficient than current ones, then switching to HPV testing might be considered beneficial for over-screened women, but this would be due to more efficient triage procedures rather than to an improved performance of the primary test.

Lastly, we did not consider a co-testing strategy, which is already recommended in the US for women aged 30-65 years.[12,17,57] Co-testing results in more screen positives than does primary HPV screening because HPV negative smears can still be cytology positive. From results of an RCT performed in the Netherlands, where women aged 30-60 years are screened every 5 years, we calculated that the number of screen positives would be 33% higher with co-testing than with primary HPV screening.[28] As a consequence, the number of screen-detected CIN 3 lesions or cancer would be 7% higher. In an RCT performed in the UK, in which women aged 20-64 years were screened with an interval of 2 to 4 years, the number of screen positives would have been 46% higher with co-testing as compared to primary HPV screening, while the number of screen-detected CIN 3 lesions would have been only 3% higher.[58] For intensively screened women, co-testing can potentially prevent slightly more cervical cancer cases than primary HPV screening, but the utility loss associated with

the additional positive screens probably outweighs these minor gains. Therefore, co-testing is expected to be even more harmful than primary HPV screening alone for over-screened women.

Conclusion

We determined the pros and cons of replacing primary cytology with primary HPV screening for women who are over-screened: i.e., from a younger age and with a shorter screening interval than recommended. Although in all scenarios more deaths would be averted by screening primarily with the HPV test, the negative effects outweighed the benefits. We may conclude that irrespective of costs, it is disputable to recommend primary HPV screening, either alone or as a co-test, as long as a substantial part of the population is still over-screened. A well-organized and structurally monitored screening program, in which primary tests taken outside the program are not reimbursed by the government, could help minimizing the number of tests taken outside the program, thereby limiting the level of over-screening.[21,59] One may consider to first further develop strategies to reduce over-screening, or at least give it high priority when issuing guidelines including primary HPV screening.

REFERENCES

1. Jemal A, Center MM, DeSantis C, et al. Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiol Biomarkers Prev 2010;19:1893-907.
2. Nanda K, McCrory DC, Myers ER, et al. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. Ann Intern Med 2000;132:810-9.
3. Bos AB, Rebolj M, Habbema JD, et al. Nonattendance is still the main limitation for the effectiveness of screening for cervical cancer in the Netherlands. Int J Cancer 2006;119:2372-5.
4. de Kok IM, van Rosmalen J, Rozemeijer K, et al. How many cervical cancer cases can potentially be prevented using a more sensitive screening test at young age? Int J Cancer 2014;134:460-6.
5. Murphy PA, Schwarz EB, Dyer JM. Cervical cancer screening practices of certified nurse-midwives in the United States. J Midwifery Womens Health 2008;53:11-8.
6. Franco EL, Rohan TE, Villa LL. Epidemiologic evidence and human papillomavirus infection as a necessary cause of cervical cancer. J Natl Cancer Inst 1999;91:506-11.
7. Walboomers JM, Jacobs MV, Manos MM, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 1999;189:12-9.
8. Arbyn M, Sasieni P, Meijer CJ, et al. Chapter 9: Clinical applications of HPV testing: a summary of meta-analyses. Vaccine 2006;24 Suppl 3:S3/78-89.
9. van Rosmalen J, de Kok IM, van Ballegooijen M. Cost-effectiveness of cervical cancer screening: cytology versus human papillomavirus DNA testing. BJOG 2012;119:699-709.
10. de Kok IM, van Rosmalen J, Dillner J, et al. Primary screening for human papillomavirus compared with cytology screening for cervical cancer in European settings: cost effectiveness analysis based on a Dutch microsimulation model. BMJ 2012;344:e670.
11. Canfell K. Progress in cancer screening: Where are we in 2014? Cancer Forum 2014;38:191-196.
12. Moyer VA, U.S. Preventive Services Task Force. Screening for cervical cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2012;156:880-91, W312.
13. Vink MA, Bogaards JA, Meijer CJ, et al. Primary human papillomavirus DNA screening for cervical cancer prevention: Can the screening interval be safely extended? Int J Cancer 2014.
14. Bell S, Porter M, Kitchener H, et al. Psychological response to cervical screening. Prev Med 1995;24:610-6.
15. Mathias JS, Gossett D, Baker DW. Use of electronic health record data to evaluate overuse of cervical cancer screening. Journal of the American Medical Informatics Association 2012;19:E96-E101.
16. Cuzick J, Clavel C, Petry KU, et al. Overview of the European and North American studies on HPV testing in primary cervical cancer screening. Int J Cancer 2006;119:1095-101.
17. Saslow D, Solomon D, Lawson HW, et al. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. Am J Clin Pathol 2012;137:516-42.
18. von Karsa L, Arbyn M, De Vuyst H, et al. European guidelines for quality assurance in cervical cancer screening. Summary of the supplements on HPV screening and vaccination. Papillomavirus Research 2015;In Press.
19. Saraiya M, Berkowitz Z, Yabroff KR, et al. Cervical cancer screening with both human papillomavirus and Papanicolaou testing vs Papanicolaou testing alone: what screening intervals are physicians recommending? Arch Intern Med 2010;170:977-85.
20. Berkowitz Z, Saraiya M, Sawaya GF. Cervical cancer screening intervals, 2006 to 2009: moving beyond annual testing. JAMA Intern Med 2013;173:922-4.

21. Arbyn M, Rebolj M, De Kok IM, et al. The challenges of organising cervical screening programmes in the 15 old member states of the European Union. Eur J Cancer 2009;45:2671-8.
22. Habbema JD, van Oortmarssen GJ, Lubbe JT, et al. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed 1985;20:79-93.
23. Statline Database. The Hague: Statistics Netherlands (CBS), 2013.
24. Nobbenhuis MA, Helmerhorst TJ, van den Brule AJ, et al. Cytological regression and clearance of high-risk human papillomavirus in women with an abnormal cervical smear. Lancet 2001;358:1782-3.
25. Anttila A, von Karsa L, Aasmaa A, et al. Cervical cancer screening policies and coverage in Europe. Eur J Cancer 2009;45:2649-58.
26. Berkhof J, Coupe VM, Bogaards JA, et al. The health and economic effects of HPV DNA screening in The Netherlands. Int J Cancer 2010;127:2147-58.
27. Casparie M, Tiebosch AT, Burger G, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol 2007;29:19-24.
28. Rijkaart DC, Berkhof J, Rozendaal L, et al. Human papillomavirus testing for the detection of high-grade cervical intraepithelial neoplasia and cancer: final results of the POBASCAM randomised controlled trial. Lancet Oncol 2012;13:78-88.
29. Bulkmans NW, Berkhof J, Bulk S, et al. High-risk HPV type-specific clearance rates in cervical screening. Br J Cancer 2007;96:1419-24.
30. Rodriguez AC, Schiffman M, Herrero R, et al. Rapid clearance of human papillomavirus and implications for clinical focus on persistent infections. J Natl Cancer Inst 2008;100:513-7.
31. Netherlands Cancer Registry (NCR). Age-specific and stage-specific cervical cancer incidence (2000-2010) and survival (1989-2009), 2012.
32. Goldie SJ, Kohli M, Grima D, et al. Projected clinical benefits and cost-effectiveness of a human papillomavirus 16/18 vaccine. J Natl Cancer Inst 2004;96:604-15.
33. Mandelblatt JS, Lawrence WF, Womack SM, et al. Benefits and costs of using HPV testing to screen for cervical cancer. JAMA 2002;287:2372-81.
34. van Ballegooijen M. Effects and Costs of Cervical Cancer Screening. Rotterdam: Erasmus University, 1998:170.
35. van den Hout WB, van der Linden YM, Steenland E, et al. Single- versus multiple-fraction radiotherapy in patients with painful bone metastases: cost-utility analysis based on a randomized trial. J Natl Cancer Inst 2003;95:222-9.
36. Siegel JE, Torrance GW, Russell LB, et al. Guidelines for pharmacoeconomic studies. Recommendations from the panel on cost effectiveness in health and medicine. Panel on cost Effectiveness in Health and Medicine. Pharmacoeconomics 1997;11:159-68.
37. Habbema D, De Kok IM, Brown ML. Cervical cancer screening in the United States and the Netherlands: a tale of two countries. Milbank Q 2012;90:5-37.
38. Bruni L, Diaz M, Castellsague X, et al. Cervical human papillomavirus prevalence in 5 continents: meta-analysis of 1 million women with normal cytological findings. J Infect Dis 2010;202:1789-99.
39. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur J Cancer 2013;49:1374-403.
40. Insinga RP, Glass AG, Myers ER, et al. Abnormal outcomes following cervical cancer screening: event duration and health utility loss. Med Decis Making 2007;27:414-22.
41. Korfage IJ, Essink-Bot ML, Westenberg SM, et al. How distressing is referral to colposcopy in cervical cancer screening?: a prospective quality of life study. Gynecol Oncol 2014;132:142-8.

42. Kitchener HC, Almonte M, Gilham C, et al. ARTISTIC: a randomised trial of human papillomavirus (HPV) testing in primary cervical screening. Health Technol Assess 2009;13:1-150, iii-iv.
43. Bulkmans NW, Rozendaal L, Snijders PJ, et al. POBASCAM, a population-based randomized controlled trial for implementation of high-risk HPV testing in cervical screening: design, methods and baseline data of 44,102 women. Int J Cancer 2004;110:94-101.
44. Ronco G, Giorgi-Rossi P, Carozzi F, et al. Results at recruitment from a randomized controlled trial comparing human papillomavirus testing alone with conventional cytology as the primary cervical cancer screening test. J Natl Cancer Inst 2008;100:492-501.
45. Naucler P, Ryd W, Tornberg S, et al. Human papillomavirus and Papanicolaou tests to screen for cervical cancer. N Engl J Med 2007;357:1589-97.
46. Ronco G, Dillner J, Elfstrom KM, et al. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. Lancet 2013.
47. Chuck A. Cost-effectiveness of 21 alternative cervical cancer screening strategies. Value Health 2010;13:169-79.
48. Cuzick J, Myers O, Hunt WC, et al. A population-based evaluation of cervical screening in the United States: 2008-2011. Cancer Epidemiol Biomarkers Prev 2014;23:765-73.
49. Rebolj M, van Ballegooijen M, Berkers LM, et al. Monitoring a national cancer prevention program: successful changes in cervical cancer screening in the Netherlands. Int J Cancer 2007;120:806-12.
50. Goldhaber-Fiebert JD, Stout NK, Salomon JA, et al. Cost-effectiveness of cervical cancer screening with human papillomavirus DNA testing and HPV-16,18 vaccination. J Natl Cancer Inst 2008;100:308-20.
51. De Vuyst H, Clifford G, Li N, et al. HPV infection in Europe. Eur J Cancer 2009;45:2632-9.
52. Dunne EF, Unger ER, Sternberg M, et al. Prevalence of HPV infection among females in the United States. JAMA 2007;297:813-9.
53. Saxena U, Sauvaget C, Sankaranarayanan R. Evidence-based screening, early diagnosis and treatment strategy of cervical cancer for national policy in low- resource countries: example of India. Asian Pac J Cancer Prev 2012;13:1699-703.
54. Sahasrabuddhe VV, Parham GP, Mwanahamuntu MH, et al. Cervical cancer prevention in low- and middle-income countries: feasible, affordable, essential. Cancer Prev Res (Phila) 2012;5:11-7.
55. Arbyn M, Kyrgiou M, Simoens C, et al. Perinatal mortality and other severe adverse pregnancy outcomes associated with treatment of cervical intraepithelial neoplasia: meta-analysis. BMJ 2008;337:a1284.
56. Jin G, LanLan Z, Li C, et al. Pregnancy outcome following loop electrosurgical excision procedure (LEEP) a systematic review and meta-analysis. Arch Gynecol Obstet 2014;289:85-99.
57. Committee on Practice Bulletins--Gynecology. ACOG Practice Bulletin Number 131: Screening for cervical cancer. Obstet Gynecol 2012;120:1222-38.
58. Kitchener HC, Gilham C, Sargent A, et al. A comparison of HPV DNA testing and liquid based cytology over three rounds of primary cervical screening: extended follow up in the ARTISTIC trial. Eur J Cancer 2011;47:864-71.
59. Elfstrom KM, Arnheim-Dahlstrom L, von Karsa L, et al. Cervical cancer screening in Europe: Quality assurance and organisation of programmes. Eur J Cancer 2015;51:950-68.

Appendix Table 1. Effects of primary cytology for 12 different screening scenarios; undiscounted numbers per 100,000 simulated women.

Screening interval	Start age	Primary screens	Positive primary screens	Referrals	False-positive referrals (no CIN detected)	CIN 1	CIN 2	CIN 3	Cervical cancer cases	Cervical cancer deaths
	30	739,525	26,923	8,868	742	2,819	2,015	3,174	398	121
5 years	25	840,937	32,721	12,139	988	3,980	2,893	4,177	358	116
	20	941,786	37,492	14,305	1,148	5,144	3,451	4,465	352	115
	30	1,118,531	38,434	11,168	1,108	4,106	2,609	3,268	332	114
3 years	25	1,315,382	47,706	15,619	1,540	5,868	3,778	4,377	269	100
	20	1,511,580	56,297	19,192	1,838	7,934	4,722	4,658	244	91
	30	1,674,942	54,320	13,581	1,609	5,634	3,110	3,172	272	96
2 years	25	1,971,111	66,941	19,192	2,259	8,003	4,489	4,238	209	82
	20	2,183,094	76,854	23,484	2,703	10,818	5,618	4,321	215	90
	30	3,346,636	99,553	18,556	3,116	9,033	3,658	2,708	222	79
1 year	25	3,858,503	118,972	25,785	4,380	12,625	5,236	3,530	176	75
	20	4,368,263	138,478	32,507	5,377	17,425	6,540	3,159	166	75

CIN = cervical intraepithelial neoplasia.

Appendix Table 2. Effects of primary HPV screening for 12 different screening scenarios; undiscounted numbers per 100,000 simulated women.

Screening interval	Start age	Primary screens	Positive primary screens	Referrals	False-positive referrals (no CIN detected)	CIN 1	CIN 2	CIN 3	Cervical cancer cases	Cervical cancer deaths
	30	740,539	35,968	11,440	1,050	4,470	2,737	3,103	283	89
5 years	25	842,392	50,462	15,881	1,485	6,290	3,933	4,113	242	84
	20	943,593	59,784	19,052	1,730	8,223	4,733	4,312	234	84
	30	1,119,883	49,809	14,100	1,550	6,353	3,261	2,883	265	95
3 years	25	1,317,518	72,465	19,951	2,294	9,053	4,729	3,847	204	82
	20	1,514,489	88,945	24,890	2,769	12,327	5,907	3,866	179	74
	30	1,676,921	66,962	16,802	2,218	8,391	3,546	2,605	231	83
2 years	25	1,974,315	98,656	23,801	3,330	11,894	5,121	3,440	170	71
	20	2,187,117	122,053	29,800	4,075	16,185	6,351	3,179	176	79
	30	3,350,355	113,824	22,033	4,209	12,146	3,519	2,124	205	72
1 year	25	3,864,616	170,288	31,145	6,426	16,955	5,036	2,719	158	68
	20	4,376,474	214,958	39,766	8,052	23,521	6,140	2,051	149	68

CIN = cervical intraepithelial neoplasia.

Chapter 4

The impact of cervical screening programs on preterm birth: a decision and cost-effectiveness analysis

Esmé I. Kamphuis*, Steffie K. Naber*, Noor A. Danhof,
J. Dik F. Habbema, Christianne J.M. de Groot, Ben W.J. Mol

Submitted.

*Shared first authorship

ABSTRACT

Importance: Cervical screening aims to reduce cervical cancer mortality. However, in younger women, treatment of cervical intraepithelial neoplasia (CIN) increases the risk of preterm birth (PTB) in later pregnancies.

Objective: To assess the impact of the start age and interval of cervical screening on the risk of PTB and subsequent neonatal outcome including the loss in neonatal quality-adjusted life years (QALYs), relative to maternal life years gained (LYG) and cost of both screening and PTB.

Design: Decision and cost-effectiveness analysis.

Setting: We compared cervical screening programs varying in age of onset (21, 24, 25, 27 or 30 years) and screening interval (3 or 5 years) in a fictive cohort of 100,000 women. We used the microsimulation screening analysis (MISCAN) model to estimate the age-specific number of cervical intraepithelial neoplasia (CIN) diagnoses, loop electrosurgical excision procedures (LEEPs), maternal LYG, QALYs gained and costs of screening. The age-specific number of LEEPs was used to calculate the number of additional PTBs, subsequent neonatal outcome, QALY loss and the associated costs.

Participants: A fictive cohort of 100,000 women.

Main outcomes: Maternal LYG and QALYs gained, additional PTB and subsequent neonatal morbidity, mortality and QALY loss and costs of both screening and PTB.

Results: Three-yearly screening from age 21 (21/3) resulted in 10,728 maternal LYG at the expense of 453 neonatal QALYs lost due to 294 additional PTBs (37 LYG per PTB), 26 additional cases of neonatal morbidity and 8 cases of mortality. Three-yearly screening from age 30 onwards (30/3) resulted in 10,419 maternal LYG at the expense of 108 neonatal QALYs loss (71 additional PTBs (147 LYG per PTB), 6 and 2 cases of neonatal morbidity and mortality). The costs of screening and subsequent PTB were €3,336 per maternal LYG for 30/3 screening and €6,336 for 21/3 screening.

Conclusion: Three-yearly cervical screening from age 30 instead of age 21 decreases the additional PTB significantly (with 76%), while slightly decreasing the maternal LYG with 2.9% and decreasing the costs per LYG by 47%. This PTB impact pleads for an individualized approach of cervical screening, taking into account the childwish of women.

INTRODUCTION

Cervical cancer screening programs aim to detect precancerous changes in the cervix that can be treated before they develop into invasive disease. Early detection and treatment of cervical intraepithelial neoplasia (CIN) have considerably reduced the incidence of cervical cancer and lowered the mortality of the disease.[1,2] In several Western countries, where screening programs have long been established, cervical cancer rates have even decreased by as much as 65% over the past 40 years.[3] Despite these successes, cervical cancer is still the fourth most common cancer among women worldwide, with an estimated 527,600 new cases and 265,700 deaths in 2012.[3] In the United States, 12,042 women were diagnosed with cervical cancer in 2012, and 4,074 women died because of the disease.[4]

Around the world, screening programs vary widely with respect to start age and screening interval.[5] In Australia routine screening with Pap smears every 2 years for women between the ages of 18 and 69 years is recommended.[6] In the United Kingdom screening starts at age 25 with a 3-year interval, and changes to a 5-year interval at age 50.[7] In the Netherlands, women are invited for cytological screening from age 30 with a 5-year interval.[8] The American College of Obstetricians and Gynaecologists (ACOG) advices to start cervical cytology screening at age 21 years with 3-year intervals. Whereas women aged 30–65 years should preferably have a Pap test and an HPV test every 5 years, they may also choose to have a Pap test alone every 3 years.[9]

In screening programs, women with abnormal cervical smears are referred to a gynaecologist for subsequent colposcopy. If high-grade changes are diagnosed, loop electrosurgical excision procedure (LEEP) is indicated according to guidelines. It has been well described that pregnant women with a history of LEEP have an increased risk of preterm birth (PTB), and associated perinatal morbidity and mortality.[10-16] As many pre-invasive lesions would regress naturally without ever resulting in cancer, i.e. up to 40% natural regression of CIN 2 lesions has been described, treatment of CIN is often unnecessary.[17,18] However, as it is not (yet) possible to differentiate between CIN lesions that will regress naturally and those that will progress to cancer, most CIN 2+ lesions are treated.

Thus, although screening programs have reduced mortality rates from cervical cancer, the diagnosis of precancerous lesions and subsequent treatment may also have resulted in unintended adverse effects due to PTB in women who became pregnant after treatment. As far as we are aware, other studies comparing the cost-effectiveness of several cervical screening programs, have not taken these unintended adverse effects into account. Therefore, the purpose of this study was to assess the impact of various screening programs on the risk of additional PTB and subsequent neonatal morbidity and mortality, including reduced quality-adjusted life years (QALY) for the child, relative to the maternal life years gained (LYG) and QALYs gained due to prevented cervical cancer and the costs of both screening and PTB.

METHODS

We used the microsimulation screening analysis (MISCAN) model to estimate the impact of eight screening programs on maternal LYG, QALYs gained and costs of screening and treatment.[19] For each screening program, we calculated the age-specific number of PTBs caused by treatment (LEEP) of precancerous lesions found during screening and estimated the subsequent neonatal morbidity and mortality resulting from these PTBs with subsequent neonatal QALYs loss and costs following from PTB.

MISCAN Model

In the MISCAN model, a study population with individual life histories is simulated. Every woman may acquire HPV infections that can progress to CIN lesions and cervical cancer. The model is based on demographic data (i.e. age-specific all-cause mortality, hysterectomy rates), natural history data (e.g. natural regression or progression of CIN lesions) and screening data (e.g. detection rates of CIN and cervical cancer).[20] The model generates age-specific outputs like detected CIN lesions, cervical cancer cases and cervical cancer deaths. We used the model to compare eight screening programs in a fictive cohort of 100,000 women in the fertile age (until 46 years), varying in age of onset of screening (21, 24, 25, 27, or 30 years of age) and interval between screening (3 or 5 years). A screening program from 21 years onwards with a three-year interval is further noted as 21/3, and similar abbreviations are used for the other programs.

Assumptions for screening and treatment

We assumed a 100% uptake of screening, and treatment with LEEP in 100% of the women with detected CIN 2 and CIN 3 lesions and in 25% of detected CIN 1 lesions (**Table 1**).

Assumptions for PTB risk after treatment and subsequent neonatal outcome

Since the fertility rate (i.e. number of children born per woman) was 1.87 in the US in 2015,[21] we assumed that all women planned to have two pregnancies. We further assumed a baseline prevalence of spontaneous PTB of 8.5% for the first pregnancy and 9.2% for the second pregnancy based on the National Vital Statistics System Birth data 2014 from the CDC.[22] As the relative risk (RR) of PTB after LEEP was assumed to be 1.39,[23] the PTB risk increased with 3.3% and 3.6% for nulliparous and primiparous women respectively (0.39 * 8.5% and 0.39 * 9.2%).

The distribution of age at which a woman gives birth to her first and second child and the distribution of PTB by gestational age (GA) separated by parity (nulliparous or primiparous) and multiplicity were also based on data from the National Vital Statistics System Birth data 2014 (**Appendix Tables 1** and **2**).[22] Neonatal morbidity and mortality probabilities due to PTB were specified by GA and type of pregnancy and are based on several randomized clinical trials in women with threatened PTB (**Appendix Tables 3** and **4**).[24]

Table 1. Assumptions for screening, treatment, and risk of preterm birth.

Parameter	Value
Screening strategy	Primary cytology, 100% uptake
Treatment with LEEP	25% of CIN 1, 100% of CIN 2 and CIN 3
Preterm birth	
Baseline risk of PTB in nulliparous pregnancies	8.5%
Baseline risk of PTB in primiparous pregnancies	9.2%
Additional PTB risk after LEEP in nulliparous pregnancies	3.3%
Additional PTB risk after LEEP in primiparous pregnancies	3.6%

CIN = cervical intraepithelial neoplasia; PTB = preterm birth; LEEP = loop electrosurgical excision procedure.

Assumptions for costs and utilities

Table 2 shows the assumptions for costs and utilities used in the analysis. Screening costs include the process used to invite women, time and travel costs required to attend screening, the test itself, cytological evaluation or HPV analysis, and registration in the screening database. We derived the costs of screening, diagnosis and treatment procedures for detected pre-invasive lesions, primary treatment of invasive cervical cancer, and treatment and palliative care for advanced cervical cancer from cost studies in the Netherlands.[25] A small (psychological) loss in quality of life was assumed for attending screening (including waiting for the result) and for being in triage (including attending follow-up screenings).[26] Larger losses in quality of life were assumed for being diagnosed and treated for CIN or cancer, and for having a terminal stage of cervical cancer.[27,28]

For each simulated woman, MISCAN can determine the state, which is classified as normal, HPV infected, having dysplasia stage CIN 1, CIN 2, CIN 3, or cervical cancer stage FIGO IA, FIGO IB, or FIGO II+. The model produces the number of life years spent in each state, numbers of certain events (e.g. screenings and cervical cancer diagnoses) in a lifetime, and the age of death. These outputs are used to determine the cost and disutility of screening and treatment for every program.

A neonatal utility loss of 1.0 was assumed in case of neonatal mortality, and smaller losses were assumed following a child alive after PTB <30 weeks (0.4) and a child alive after PTB between 30 and 37 weeks of GA (0.25).[29] Based on the current US life expectancy of 78 and a discount rate of 3%, the neonatal QALY loss due to PTB was calculated.

The costs resulting from PTB were specified by GA and type of pregnancy based on data collected in several randomized clinical trials in women with threatened PTB (**Appendix Table 4**).[24]

Statistical analyses

For every screening program, the MISCAN model estimates the amount of CIN diagnoses, maternal LYG and QALYs gained and costs per 100,000 simulated women, as compared to the situation without screening. The additional number of PTBs was calculated by multiply-

Table 2. Assumptions on costs and utilities.

Parameter	Value
Costs of screening and treatment	
Invitation letter	€4.85
Primary cytology screening	€66.09
Reflex HPV triage (after 6 months only)	€29.00
Cytology triage after 6 or 18 months	€63.59
Diagnosis and treatment of pre-invasive stages	
False positive	€296
CIN 1	€924
CIN 2	€1,368
CIN 3	€1,602
Diagnosis and treatment of cancer	
FIGO IA	€5,246
FIGO IB	€12,440
FIGO II + (detected by screening)	€12,261
FIGO II + (detected by symptoms)	€11,451
Terminal care	€27,859
Neonatal utility loss	
Neonatal mortality	1.0
Morbidity PTB <30 weeks	0.4
Morbidity PTB 30-37 weeks	0.25

PTB = preterm birth; CIN = cervical intraepithelial neoplasia; FIGO = International Federation of Gynecology and Obstetrics; HPV = human papillomavirus.

ing the number of women with a LEEP with the additional probability of PTB after LEEP based on age and parity. For example, from **Appendix Table 1** it derives that 17.2% (5.9% + 5.8% + 5.5%) of the American women would give birth to their first child at age 21, 22, or 23. Of these women, 1,583 in the cohort will have undergone LEEP prior to their pregnancy in case of 21/3 screening program. With an additional PTB risk after LEEP of 3.3% as compared to the baseline risk of 8.5% (**Table 1**), this leads to 9.0 (17.2% * 1583 * 3.3% = 9.0) additional PTBs. Adding all the PTBs for first and second pregnancy for all ages between 21-46 years generates the total amount of additional PTBs due to that screening program.

Subsequently, based on the number of PTBs, we calculated the neonatal morbidity and mortality in relation to GA and parity. For example, from the assumptions in **Appendix Table 2**, it derives that 31.5% of all PTBs in the first pregnancy are singletons born between 36-37 weeks of GA. Multiplying this by the 0.5% probability of morbidity of a singleton born between 36-37 weeks of GA (**Appendix Table 3**) and the number of PTBs in a screening program, this generates an expected amount of neonatal morbidity due to additional PTB due to a screening program for singletons from first pregnancy at this GA. Adding up these

numbers for all the expected PTBs at different GA leads to the expected total additional neonatal morbidity for the screening program.

Similarly we calculated the additional neonatal mortality and costs of PTB for each screening program. Finally, the neonatal morbidity and mortality rates were translated in the number of reduced QALYs for the child.

We also calculated the costs per absolute QALY gained (maternal QALYs gained minus neonatal QALYs lost). Finally, we calculated the ratio LYG per additional PTB to compare the impact of several screening programs.

RESULTS

Maternal outcome

Table 3 shows the impact of the eight screening programs on the amount of maternal LYG, maternal QALYs gained and cost of both screening and treatment. The amount of LYG varied from 9,809 (24/5) to 10,728 (21/3) and were lower for screening programs with a 5-year interval (9,809-10,379) compared to a 3-year interval (10,419-10,728).

The maternal QALYs gained ranged from 9,759 (24/5) to 10,621 (27/3), again being lower for a 5-year screening interval (9,759-10,316) than for a 3-year interval (10,482-10,621).

Although the maternal LYG are highest in the most intensive screening program (21/3), the maternal QALYs gained from the 30/3 program were higher than from the 21/3 program due to maternal morbidity created by the intensive screening and treatment itself.

The total costs varied from € 18.7 million (30/5) to € 62.7 million (21/3) due to more or less intensive screening.

Table 3. Maternal outcome per 100,000 women.

Program	LYG	QALYs gained	Costs[a]
21/3[b]	10,728	10,482	€ 62.7
24/3	10,700	10,558	€ 53.3
27/3	10,628	10,621	€ 43.2
30/3	10,419	10,533	€ 33.5
21/5	10,379	10,316	€ 38.0
24/5	9,809	9,759	€ 30.5
25/5	10,086	10,094	€ 29.3
30/5	9,839	10,005	€ 18.7

LYG = life years gained; QALY = quality-adjusted life years.

[a] Costs of screening and treatment in million euros (rounded).

[b] Start age of screening/screening interval

Neonatal outcome

Table 4 shows the impact of the eight screening programs on the neonatal outcome, i.e., the number of additional PTBs, neonatal morbidity and mortality, neonatal loss in QALYs, and finally the costs of PTB.

Table 4. Neonatal outcome per 100,000 women

Program	Additional PTB	Morbidity	Mortality	QALYs loss	Costs[a]
21/3[b]	294	26	8	453	€5.3
24/3	226	20	6	347	€4.0
27/3	138	12	4	212	€2.5
30/3	071	6	2	108	€1.3
21/5	226	20	6	348	€4.1
24/5	181	16	5	278	€3.2
25/5	159	14	4	244	€2.8
30/5	59	5	2	90	€1.1

PTB = preterm birth; QALY = quality-adjusted life years.

[a] Costs of preterm birth in million Euros (rounded).

[b] Start age of screening/screening interval

The additional PTBs ranged from 59 (30/5) to 294 (21/3), with subsequent neonatal morbidity and mortality ranging from 5 (30/5) to 26 (21/3) and from 2 (30/5) to 8 (21/3), respectively. The additional PTB rates decrease with a later onset of screening. The neonatal loss in QALYs varied from 90 for the least intensive (30/5) to 453 for the most intensive (21/3) screening program respectively, with costs varying from €1.1 to €5.3 million. Both neonatal QALY loss and costs decrease with a later onset of screening.

Combined neonatal and maternal outcome

Table 5 shows the overall impact, i.e. the combined neonatal and maternal outcome. The amount of maternal LYG per additional PTB varied from 36.5 to 166.7 due to the most (21/3) or least (30/5) intense screening program respectively, with subsequent total costs per maternal LYG of € 6,336 and € 2,010 respectively.

The overall QALYs gained (i.e. maternal QALYs gained minus neonatal QALYs lost) varied from 9,482 (24/5) to 10,425 (30/3) and were lower for screening programs with a 5-year interval (9,482-9,968) compared to a 3-year interval (10,028-10,425). The expense per QALY gained ranged from €1,995 (30/5) to €6,778 (21/3).

Three-yearly screening from age 30 instead of age 21 involves slightly more maternal QALYs gained (10,533 instead of 10,482 per 100,000 women; Table 3), and is associated with 76% fewer additional PTBs due to cervical screening (294 instead of 71 per 100,000 women; Table 4). By taking into account also the neonatal QALYs lost due to these PTBs, overall QALYs will increase with 2.9% when increasing the start age of 3-yearly screening from 21 to 30.

Table 5. Combined maternal and neonatal outcome per 100,000 women

Program	LYG / PTB	Total costs[a] / LYG	Total QALYs[b]	Total Costs[a] / Total QALYs[b]
21/3[c]	36.5	€6,336	10,028	€6,778
24/3	47.4	€5,356	10,211	€5,613
27/3	76.8	€4,295	10,409	€4,385
30/3	147.0	€3,336	10,425	€3,334
21/5	45.9	€4,050	9,968	€4,217
24/5	54.3	€3,443	9,482	€3,562
25/5	63.5	€3,187	9,850	€3,263
30/5	166.7	€2,010	9,914	€1,995

LYG = life years gained; PTB = preterm birth; QALYs = quality-adjusted life years.

[a] Total costs include the costs of cervical screening and treatment, and of PTB.

[b] Total QALYs reflects the gain in maternal QALYs minus the loss in neonatal QALYs

[c] Start age of screening/screening interval

As costs are also higher for screening from age 21, the cost per QALY gained will decrease from €6,778 to €3,334.

DISCUSSION

We assessed the risk of additional PTB and subsequent neonatal morbidity, mortality and associated QALY loss due to to treatment of CIN in relation to maternal LYG and QALYs gained and costs for different screening programs and the additional PTBs. We found that both the age of onset of screening and the screening interval affect the risk of PTB. Screening for cervical cancer with a 3-year interval from age 30 instead of age 21 decreased additional PTB sand subsequent neonatal QALY loss with 76% for a slight decrease in maternal LYG while decreasing the costs per overall QALY gained by 50%.

Strength and limitations

A strength of our analysis is the fact that we model eight different programs for screening, varying both the age of onset of screening and the screening interval, to gain separate insight into these two variables. Furthermore, we analyzed the costs for the different programs, not only those related to the screening, diagnosis and treatment, but also for the additional PTBs. Therefore taking into account the additional costs as a consequence from the intensity of screening programs leading to additional PTBs. A third strength of our analysis is that we used accurate data for both PTB rates and distribution of GA separated by parity and multiplicity and distribution of age at which women give birth to their children, which are representative for the US situation in 2014, the most recent available data. This leads to a realistic view on the impact of several screening programs, taking into account

the relatively young age at which women give birth to their first and second child in the US and the relatively high PTB rates. Finally we are unaware of other studies comparing several screening programs in terms of the risk of PTB and subsequent neonatal QALY loss as a consequence of cervical screening, relative to costs and maternal QALYs gained.

The study also has limitations. As the MISCAN model is based on Dutch characteristics, it is not fully representative for the US population. For example, the US cervical cancer incidence and mortality rate is slightly higher and screening could therefore result in more life years gained than in the Netherlands. However, although absolute rates might change slightly, proportions between the programs would not alter. Since we used this model as a proof of principle to demonstrate the differences in impact of various screening programs, the chance that absolute numbers would alter our conclusions is unlikely. Moreover, it is unlikely for model assumptions regarding the natural history of the disease to differ much across countries.

In our analysis, we presumed an uptake of cervical screening of 100%. Although we acknowledge that in real life this will not be the case, we did so to show what the potential consequences would be for a woman who considers screening and has a potential child-wish later in life.

Interpretation

Although there is no doubt about the success of cervical cancer screening programs, there is debate on the appropriate age to start screening, as is demonstrated by the wide range of screening programs implemented over the world varying in age of onset from 18 to30 years. Given the high prevalence of regressive lesions at young age, screening at young age could lead to a high treatment rate, thereby potentially influencing the outcomes of future pregnancies. This issue makes the question at what age to start screening for cervical cancer more pressing. In fact, when taking into account the morbidity generated by screening and subsequent treatment and translating that to QALYs, we already found that a 21/3 strategy did not add to a 30/3 strategy. However, even when considering maternal LYG as outcome measure (i.e. not considering any potential loss in quality of life due to screening and treatment), the merit of early screening is doubtful.

Several systematic reviews and meta-analyses have been performed on the risk of PTB following different kinds of treatment for CIN and showed an increased risk of PTB.[11,12] A recent meta-analysis, comparing the impact of LEEP on PTB to women with CIN without LEEP, did not find a significant difference and suggested the association to be confounded by CIN underlying the surgery, or by infection leading to CIN.[16] However, another recent meta-analysis, including 4 additional studies, compared the risk of PTB in women who had had cervical surgery for CIN to women who had CIN but had not been treated, confirmed an increased risk after cervical surgery (RR of 1.67, 95% CI: 1.04 to 2.67), which was partly attributed to treatment for CIN during pregnancy.[23] A subgroup analysis of treatment before pregnancy also showed an increased risk, although insignificant (RR of 1.39, 95% CI: 0.85 to 2.04).

Castanon et al. assessed the impact of increasing depth or volume of the excision on PTB in women with cervical dysplasia.[14] They reported that the risk of PTB doubles with larger excisions, especially over 15 mm or 2.66cm^3 as compared to a small LEEP. Furthermore they report that this risk does not decrease with increasing time from excision to conception. Castanon et al. also report that the increased risk of PTB is not restricted to the first pregnancy post treatment but remains for the second and subsequent pregnancies.[15] The difference in PTB rate between less and more invasive cervical surgery or time between excision and pregnancy is unlikely to be confounded. Thus, the present analysis shows that in women who might want to conceive in the future, it is important to limit the depth of excisions as much as possible. Cold coagulation could be an alternative to cervical excision that is associated with a lower risk of preterm birth.[30] Determining the exact risk goes beyond the scope of this manuscript, however in view of the current knowledge, we believe that the results of our analysis would plead for a general reticence with cervical screening and treatment in women below 30. This is important because in the US alone, more than 400,000 women are diagnosed with CIN annually, the majority at reproductive age.[31]

Our assumptions on prevalence of HPV and abnormal cervical histology are likely to be influenced by HPV vaccination programs. As these programs start to have an effect, the number of screen positive women and the number of women requiring treatment might decrease, as well as the resulting number of PTB. However, if the distribution of the type of cervical lesions does not change, then the balance between maternal LYG and risk of PTB is expected to remain similar.

The final choice might be influenced by the presence of other risk factors for cervical disease, such as young age at first intercourse, high parity, long-term use of oral contraceptives and smoking.[32-35] As an alternative to surgery, a woman who wants to have children and who has an abnormal Pap smear, could opt for more frequent follow-up to monitor progression or regression of disease, instead of immediate surgical excision. Also, women with abnormal cytology could decide to have their children earlier, and postpone further diagnosis until after pregnancy. Finally, deep and thick excisions should be reconsidered. The final decision for screening and subsequent treatment should be made in a process of shared decision making.

In summary, we found that both the age of onset of screening and the screening interval of a cervical screening program impact the additional risk of PTB due to cervical surgery. Our analysis pleads for tailored management of cervical screening results in women of childbearing age, in which one, based on the findings of cervical screening and a woman's risk profile, should consider less invasive treatment than the standard regime.

REFERENCES

1. Habbema D, De Kok IM, Brown ML. Cervical cancer screening in the United States and the Netherlands: a tale of two countries. Milbank Q. 2012 Mar; 90(1): 5-37.
2. Edwards BK, Noone AM, Mariotto AB et al. Annual Report to the Nation on the status of cancer, 1975-2010, featuring prevalence of comorbidity and impact on survival among persons with lung, colorectal, breast, or prostate cancer. Cancer. 2014 May 1; 120(9):1290-314.
3. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015 Mar; 65(2): 87-108.
4. U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2012 Incidence and Mortality Web-based Report. Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute; 2015.
5. Anttila A, von Karsa L, Aasmaa A, et al. Cervical cancer screening policies and coverage in Europe. Eur J Cancer. 2009 Oct; 45(15):2649-58.
6. Australian Goverment Department of Health. National Cervical Screening Program. http://www.cancerscreening.gov.au/internet/screening/publishing.nsf/Content/about-the-programme-2. Accessed March 2, 2017.
7. National Health Service Cervical Screening Program. About cervical cancer screening. Sheffield, UK: National Health Service; 2011. www.cancerscreening.nhs.uk/cervical/about-cervical-screening.html. Accessed May 11, 2014.
8. Van Ballegooijen M, Hermens R. Cervical cancer screening in The Netherlands. Eur J Cancer. 2000 Nov; 36(17): 2244-6.
9. American Congress of Obstetricians and Gynecologists. ACOG Practice Bulletin No. 109: Cervical cytology screening. Obstet. Gynecol., 114 (6) (2009), pp. 1409–1420.
10. Kyrgiou M, Koliopoulos G, Martin-Hirsch P, Arbyn M, Prendiville W, Paraskevaidis E. Obstetric outcomes after conservative treatment for intraepithelial or early invasive cervical lesions: systematic review and meta-analysis. Lancet. 2006 Feb 11;367(9509):489-98.
11. Bruinsma FJ, Quinn MA. The risk of preterm birth following treatment for precancerous changes in the cervix: a systematic review and meta-analysis. BJOG 2011;118:1031–41.
12. Arbyn M, Kyrgiou M, Simoens C, Raifu AO, Koliopoulos G, Martin-Hirsch P, et al. Perinatal mortality and other severe adverse pregnancy outcomes associated with treatment of cervical intraepithelial neoplasia: meta-analysis. BMJ 2008; 337: a1284.
13. Miller ES, Sakowicz A, Grobman WA. The association between cervical dysplasia, a short cervix, and preterm birth. Am J Obstet Gynecol. 2015 Oct; 213 (4): 543.e1-4.
14. Castanon A, Landy R, Brocklehurst P, et al. PaCT Study Group. Risk of preterm delivery with increasing depth of excision for cervical intraepithelial neoplasia in England: nested case-control study. BMJ. 2014 Nov 5;349:g6223.
15. Castañon A, Landy R, Brocklehurst P, et al. PaCT Study Group. Is the increased risk of preterm birth following excision for cervical intraepithelial neoplasia restricted to the first birth post treatment? BJOG. 2015 Aug;122(9):1191-9.
16. Conner SN, Frey HA, Cahill AG, Macones GA, Colditz GA, Tuuli MG. Loop Electrosurgical Excision Procedure and Risk of Preterm birth: A systematic Review and Meta-analysis. Obstetrics & Gynaecology 2014 Apr; 123(4): 752-761.
17. Vesco KK, Whitlock EP, Eder M, Burda BU, Senger CA, Lutz K. Risk factors and other epidemiologic considerations for cervical cancer screening: a narrative review for the U.S. Preventive Services Task Force. Ann Intern Med. 2011 Nov 15; 155(10):698-705, W216.

18. Castle PE, Schiffman M, Wheeler CM, Solomon D. Evidence for frequent regression of cervical intraepithelial neoplasia-grade 2. Obstet Gynecol. 2009; 113: 18-25.
19. Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed 1985; 20: 79-93.
20. de Kok IM, van Rosmalen J, van Ballegooijen M. Description of MISCAN – cervix. Web appendix accompanying 'Primary screening for human papillomavirus compared with cytology screening for cervical cancer in European settings: Cost effectiveness analysis based on a Dutch microsimulation model'. BMJ 2012 Mar 5; 344: e670.
21. Central Intelligence Agency. The World Factbook. https://www.cia.gov/library/Publications/the-world-factbook/rankorder/2127rank.html. Accessed March 3, 2017.
22. Centers for Disease Control and Prevention. About Natality, 2007-2015. https://wonder.cdc.gov/natality-current.html. Accessed March 3, 2017.
23. Danhof NA, Kamphuis EI, Limpens J, van Lonkhuijzen LRCW, Pajkrt E, Mol BWJ. The risk of preterm birth of treated versus untreated cervical intraepithelial neoplasia (CIN): a systematic review and meta-analysis. Eur J Obstet Gynecol Reprod Biol. 2015 May;188:24-33.
24. GJ van Baaren, MJCS Peelen, E Schuit, et al. Preterm birth in singleton and multiple pregnancies: evaluation of costs and perinatal outcomes. Eur J Obstet Gynecol Reprod Biol. 2015 Mar; 186: 34-41.
25. Van Ballegooijen M, Rebolj M, Essink-Bot ML, Meerding WJ, Berkers LM, Habbema JDF. De effecten en kosten van het bevolkingsonderzoek naar baarmoederhalskanker in Nederland na de herstructurering. http://www.rivm.nl/dsresource?objectid=bfb9ef32-ebe3-4468-8df2-f258bfd0fb61&type=org&disposition=inline. Accessed March 3, 2017.
26. Van Ballegooijen M. Effects and Costs of Cervical Cancer Screening. Rotterdam, the Netherlands: Erasmus University; 1998. http://hdl.handle.net/1765/16983. Accessed July 10, 2014.
27. Mandelblatt JS1, Lawrence WF, Womack SM, et al. Benefits and costs of using HPV testing to screen for cervical cancer. JAMA. 2002 May 8;287(18):2372-81.
28. Goldie SJ, Kohli M, Grima D, et al. Projected clinical benefits and cost-effectiveness of a human papillomavirus 16/18 vaccine. J Natl Cancer Inst. 2004 Apr 21;96(8):604-15
29. Tengs TO, Wallace A. One thousand health-related quality-of-life estimates. Med Care. 2000 Jun;38(6):583-637.
30. Dolman L, Sauvaget C, Muwonge R, Sankaranarayanan R. Meta-analysis of the efficacy of cold coagulation as a treatment method for cervical intraepithelial neoplasia: a systematic review. BJOG. 2014 Jul;121(8):929-42.
31. Henk HJ, Insinga RP, Singhal PK, Darkow T. Incidence and costs of cervical intraepithelial neoplasia in a US commercially insured population. J Low Genit Tract Dis 2010;14:29–36.
32. International Collaboration of Epidemiological Studies of Cervical Cancer. Comparison of risk factors for invasive squamous cell carcinoma and adenocarcinoma of the cervix: collaborative reanalysis of individual data on 8,097 women with squamous cell carcinoma and 1,374 women with adenocarcinoma from 12 epidemiological studies. Int J Cancer. 2007; 120:885-91.
33. Solomon D, Davey D, Kurman R, et al. Forum Group Members. The 2001 Bethesda System: terminology for reporting results of cervical cytology. JAMA. 2002; 287:2114-9.
34. Castellsagué X, Diaz M, de Sanjosé S, et al. International Agency for Research on Cancer Multicenter Cervical Cancer Study Group. Worldwide human papillomavirus etiology of cervical adenocarcinoma and its cofactors: implications for screening and prevention. J Natl Cancer Inst. 2006; 98:303-15
35. International Collaboration of Epidemiological Studies of Cervical Cancer, Appleby P, Beral V, Berrington de González A, et al. Carcinoma of the cervix and tobacco smoking: collaborative reanalysis of individual data on 13,541 women with carcinoma of the cervix and 23,017 women without carcinoma of the cervix from 23 epidemiological studies. Int J Cancer. 2006; 118:1481-95.

Appendix Table 1. Age distribution of pregnancy by parity.[22]

	Nulliparous pregnancies		Primiparous pregnancies	
Age	**Number**	**Percentage**	**Number**	**Percentage**
<13	81	0.0%	0	0.0%
13	381	0.0%	1	0.0%
14	2,270	0.2%	26	0.0%
15	7,805	0.5%	231	0.0%
16	18,978	1.2%	1,022	0.1%
17	34,513	2.2%	3,567	0.3%
18	57,753	3.7%	10,124	0.8%
19	86,698	5.6%	21,812	1.7%
20	95,801	6.2%	37,012	2.9%
21	92,101	5.9%	49,875	3.9%
22	89,462	5.8%	62,014	4.9%
23	84,925	5.5%	67,975	5.4%
24	83,234	5.4%	71,286	5.6%
25	82,469	5.3%	71,997	5.7%
26	83,686	5.4%	72,948	5.8%
27	86,020	5.5%	74,470	5.9%
28	88,770	5.7%	77,737	6.1%
29	87,817	5.7%	80,079	6.3%
30	83,146	5.4%	79,869	6.3%
31	76,856	5.0%	80,807	6.4%
32	66,475	4.3%	76,781	6.1%
33	54,586	3.5%	69,404	5.5%
34	45,328	2.9%	61,236	4.8%
35	36,322	2.3%	50,918	4.0%
36	28,295	1.8%	40,510	3.2%
37	21,654	1.4%	31,686	2.5%
38	16,179	1.0%	23,692	1.9%
39	12,522	0.8%	17,850	1.4%
40	8,940	0.6%	12,159	1.0%
41	6,282	0.4%	8,080	0.6%
42	4,328	0.3%	5,213	0.4%
43	2,716	0.2%	3,121	0.3%
44	1,675	0.1%	1,750	0.1%
45	941	0.06%	923	0.07%
46	941	0.06%	473	0.04%
47	941	0.06%	261	0.02%
48	204	0.01%	134	0.01%
49	139	0.01%	111	0.01%
>49	238	0.02%	180	0.01%
Total	1,551,472	100.0%	1,267,334	100.0%

Appendix Table 2. Distribution of preterm births by gestational age and singleton or multiple pregnancy, for nulliparous and primiparous women.[22]

	Nulliparous women		Primiparous women	
GA (weeks)	**Singletons**	**Multiples**	**Singletons**	**Multiples**
24	1.2%	0.2%	0.6%	0.3%
25	1.4%	0.3%	0.8%	0.3%
26	1.5%	0.3%	0.8%	0.4%
27	1.6%	0.3%	1.0%	0.5%
28	1.8%	0.4%	1.2%	0.6%
29	2.0%	0.5%	1.4%	0.7%
30	2.5%	0.6%	1.7%	0.8%
31	3.0%	0.9%	2.2%	1.2%
32	4.3%	1.4%	3.2%	1.9%
33	6.1%	1.7%	4.6%	2.7%
34	10.3%	2.8%	8.9%	4.3%
35	16.0%	3.1%	14.6%	5.4%
36	31.5%	4.3%	32.4%	7.5%
Total	83.2%	16.8%	73.5%	26.5%

GA = gestational age.

Appendix Table 3. Probability of neonatal morbidity and mortality by gestational age and singleton or multiple pregnancy.[24]

	Singletons		Multiples	
GA (weeks)	**Morbidity**[a]	**Mortality**	**Morbidity**[a]	**Mortality**
24	50%	50%	22%	75%
25	54%	39%	37%	47%
26	58%	24%	57%	22%
27	54%	9%	60%	13%
28	25%	5%	54%	11%
29	34%	4%	39%	12%
30	26%	3%	30%	7%
31	19%	3%	29%	5%
32	11%	2%	22%	4%
33	6%	2%	15%	4%
34	2%	1%	10%	2%
35	1%	0%	7%	1%
36	0.5%	0%	4%	1%

GA = gestational age.

[a] Morbidity is defined as chronic lung disease (in need of oxygen at 28 days after birth or clinically determined bronchopulmonary dysplasia), intraventricular haemorrhage ≥ grade 2, periventricular leukomalacia ≥ grade 1, proven sepsis or necrotising enterocolitis.

Appendix Table 4. Costs per PTB by GA and singleton or multiple pregnancy.[24]

GA (weeks)	Singletons	Multiples
24	€113,432	€77,726
25	€62,671	€155,315
26	€75,218	€191,753
27	€77,575	€160,232
28	€60,158	€156,729
29	€44,615	€88,292
30	€31,559	€61,121
31	€27,357	€30,625
32	€19,973	€41,468
33	€16,142	€29,880
34	€11,622	€19,040
35	€5,903	€15,451
36	€1,952	€8,426

PTB = preterm birth, GA = gestational age.

PART III

Increasing Benefits

Chapter 5

Cervical cancer incidence after a normal cytological sample in routine screening using SurePath, ThinPrep and conventional cytology: a population-based study

Kirsten Rozemeijer, Steffie K. Naber, Corine Penning, Lucy I.H. Overbeek, Caspar W.N. Looman, Inge M.C.M. de Kok, Suzette M. Matthijsse, Matejka Rebolj, Folkert J. van Kemenade, Marjolein van Ballegooijen

BMJ. 2017;356:j504

ABSTRACT

Objective: To compare the cumulative incidence of cervical cancer diagnosed within 72 months after a normal screening sample between conventional cytology and liquid-based cytology tests SurePath and ThinPrep.

Design: Retrospective population-based cohort study.

Setting: Nationwide network and registry of histo- and cytopathology in the Netherlands (PALGA), January 2000 to March 2013.

Population: Women with 5,924,474 normal screening samples (23,833,123 person years).

Exposure: Use of SurePath or ThinPrep versus conventional cytology as screening test.

Main outcome measure: The 72-month cumulative incidence of invasive cervical cancer after a normal screening sample for each screening test. Cox regression analyses assessed the hazard ratios (HRs), adjusted for calendar time, age, screening history, and socioeconomic status and including laboratories as random effects.

Results: The 72-month cumulative cancer incidence was 58.5 (95% confidence interval (CI): 54.6 to 62.7) per 100,000 normal conventional cytology samples, compared with 66.8 (95% CI: 56.7 to 78.7) for ThinPrep and 44.6 (95% CI: 37.8 to 52.6) for SurePath. Compared with conventional cytology, the hazard of invasive cancer was 19% lower [HR of 0.81 (95% CI: 0.66 to 0.99)] for SurePath, mainly caused by a 27% lower hazard [HR of 0.73 (95% CI: 0.57 to 0.93)] of a clinically detected cancer. For ThinPrep, the hazard was on average 15% higher [HR of 1.15 (95% CI: 0.95 to 1.38)], mainly caused by a 56% higher hazard of a screen-detected cancer [HR of 1.56 (95% CI: 1.17 to 2.08)].

Conclusions: These findings should provoke reconsideration of the assumed similarity in sensitivity to detect progressive cervical intraepithelial neoplasia between different types of LBC and conventional cytology.

INTRODUCTION

The use of conventional cytology as the primary test method has been replaced by the use of liquid-based cytology (LBC) in many countries with organized cervical cancer screening programs, such as the UK, the Netherlands, and Denmark.[1,2] The main advantages of using LBC instead of primary conventional cytology are facilitation of reflex testing (i.e. the residual material can be tested for the presence of the human papillomavirus (HPV) in case of borderline/mildly dyskaryotic smears)[3,4] and reduction in the number of slides of unsatisfactory quality.[5-9] The sensitivity of LBC for detecting cervical intraepithelial neoplasia (CIN) 2+ lesions is believed to be similar to that of conventional cytology.[10,11] However, the literature has been dominated by many studies comparing CIN detection between ThinPrep and conventional cytology.[12-17] whereas only two studies have compared CIN detection between SurePath and conventional cytology.[7,18] Therefore, we compared CIN 2+ detection rates between these three types of cytology tests in our previous study, including more than six million smears taken within the Dutch cervical cancer screening program.[19] Whereas the use of SurePath led to an 8% increase in detection of CIN 2+ compared with conventional cytology, the use of ThinPrep did not affect CIN 2+ detection rates. These results were compatible with the results of other studies.[12-16,20]

Detecting more CIN as a result of an abnormal screening test is expected to deplete the pool of lesions that would have progressed to cancer. However, in the absence of screening (and associated treatment) only a fraction of screen-detected CIN would progress to cervical cancer (i.e. the progressive CIN lesions), so detecting more CIN lesions is not always equivalent to preventing more cervical cancers. To assess whether the ability to detect progressive CIN lesions differs between the types of screening tests, the probability of a diagnosis of invasive cervical cancer in the period after a normal test result (i.e. the progressive CIN that screening has missed) has to be compared. As the incidence of invasive cervical cancer after a normal test result is low (6-year cumulative incidence rate: 48 per 100,000 normal smears (95% confidence interval (CI): 43 to 54)),[21] such a comparison can be made only by using an observational population-based study in which a large number of samples can be evaluated.

In the Netherlands, organized cervical cancer screening has existed since the 1980s, and women aged 30 to 60 years have been invited every five years since 1996. Until 2016, the screening strategy consisted of primary cytology screening with cytology triage testing, the latter either alone or in combination with HPV testing. All cervix uteri cytological and histological tests taken inside and outside the Dutch screening program are registered in the nationwide network and registry of histo- and cytopathology in the Netherlands (PALGA).[22] By using these data, we assessed the cumulative incidence of invasive cervical cancer detected within 72 months after a normal screening test result (i.e. diagnosed in the next screening round or outside the organized Dutch screening program). If this incidence varies among the three tests (SurePath, ThinPrep, and conventional cytology), the sensitivity for progressive CIN lesions probably differs between them. In addition, we stratified for the

reason for cervical examination that led to the detection of cervical cancer (screen-detected when program smear detected, or clinically detected in all other cases, which includes opportunistic screening as well as direct biopsies).

METHODS

Information on all cytological and histological examinations of the cervix uteri taken in the Netherlands between January 2000 and March 2013 were available and retrieved from the national PALGA database. Multiple quality checks ensured the reliability of the retrieved data.[23,24] We identified women through their birth date and the first eight letters of their (maiden) family name. This identification code enables linkage of multiple tests belonging to the same woman, allowing us to follow individual screening histories.

We identified and selected episodes starting with a normal primary screening sample taken within the Dutch screening program between January 2000 and March 2012. We identified screening samples through the reason for taking the sample being participation in the program, which is routinely registered in PALGA. We also selected women with a primary sample of inadequate quality followed by a normal sample within the same episode. We defined an episode as starting with a primary test followed by one or more secondary tests in case the result was abnormal (at least borderline or mild dyskaryosis) or of inadequate quality. Unless the follow-up of a primary test had already been completed according to the guidelines (e.g. by two consecutive normal samples after a screening result with borderline dyskaryosis), we conidered tests taken within four years after a primary test to be secondary tests.[25] We labeled all other tests as primary tests.

We stratified normal primary screening samples by the type of cytology test used (SurePath, ThinPrep, or conventional cytology). As PALGA does not register this routinely, regional coordinating pathologists obtained conversion dates (fixed to the first date of the yearly quarter) from the individual laboratories located in their region. In the Netherlands, it is standard practice for laboratories to supply general practitioners with cytology kits and thereby determine the type of cytology test that is used. We linked these conversion dates to the laboratory involved and the examination date as a proxy for which type of primary cytology test was used.

Follow-up ended at the date of the primary test of the next episode, which resulted in a cervical cancer diagnosis (a case) or not, on 31 March 2013 (the end of the database), or on completion of the 72-month period, whichever came first. We censored the follow-up at 72 months because it covers the invitation for the next screening round, which takes place 60 months after a normal screening sample. By definition, after a normal primary screening test, all new tests give rise to a new episode regardless of the reason for taking the test and the type of test. We identified histologically confirmed cases of cervical cancer by selecting all PALGA records that included pathology codes describing invasive cancers originating in

the cervix uteri. These codes were manually checked to avoid over-counting both of non-invasive lesions and primary cancers originating elsewhere.

As women in the Netherlands are invited for screening in the year they turn 30, 35, 40, 45, 50, 55, and 60, we categorized age as: 29-33, 34-38, 39-43, 44-48, 49-53, 54-58, and 59-63 years at the time of the normal primary cytological sample. We also defined calendar year at the time of the normal cytological sample. We defined socioeconomic status (SES), categorized as low, middle, or high, by the status score. This is an ecological variable based on the household characteristics of the four-digit postcode area where the woman was living at the time of the primary test.[26] Status scores per four-digit postal code came from the Netherlands Institute for Social Research and were based on mean income, percentage of households with a low income, percentage of households with (on average) a low education, and unemployment rate in 2010. Low SES corresponded to a status score lower than -1 (i.e. average status score minus 1 standard deviation), intermediate SES to a score between -1 and 1, and high SES to a score higher than 1 (i.e. average status score plus 1 standard deviation). We categorized screening history as no history of cytological smears (inside or outside the screening program) before the normal screening sample, one cytological smear that was taken less than seven years before the normal screening sample, one cytological smear that was taken more than seven years before the normal screening sample, at least two cytological smears with the last being taken less than seven years before the normal screening sample, and at least two cytological smears with the last being taken more than seven years before the normal screening sample.

Statistical analyses

Laboratories implemented LBC testing at different points in time. Therefore, we expected follow-upand calendar time to differ between the three types of cytology tests. As demographic characteristics of screened women (age, screening history, and SES) probably differed between laboratories, we expected them to differ between the types of cytology tests as well. As age, SES, screening history, and calendar time were all associated with CIN and/or cervical cancer detection rates,[27-30] they were all potential confounding factors. We used a Pearson's chi-squared test to test whether their distributions differed between the types of cytological tests. We considered a p value of less than 0.05 to be statistically significant. Missing values were imputed with 10 imputation sets.

Cumulative incidence and hazard ratio

For each type of cytology test, we did a Kaplan-Meier analysis to calculate the cumulative incidence of invasive cervical cancer per 100,000 normal cytological screening samples. We took differences in follow-up time into account and estimated the 95% CIs by non-parametric Kaplan-Meier product-limit estimator for log(hazard).[21,31] We used the R package "coxme"[32] to do multilevel Cox regression analyses to compare the hazard of cervical cancer between the types of cytology tests, taking differences in follow-up time into account and adjusting for the confounding factors calendar time, SES, age, and screening history. We

included the determinant laboratory as a random effect in the model to take account of clustering of the data at the laboratory level. In addition, we stratified for the reason for the cervical examination that led to the cervical cancer diagnosis (i.e. screen-detected when detected by a program smear, or clinically detected in all other cases, including opportunistic screening as well as direct biopsies). We tested time dependencies of the hazard ratios (HRs) statistically by splitting the total follow-up time into two periods with a roughly equal number of cases. Subsequently, we assessed HRs for each time period. If the sum of the deviance of both sub-models was significantly lower than the deviance of the original model, we considered the HR to be time dependent as it differed significantly between the time periods.

Sensitivity analyses

In the first sensitivity analysis, we restricted our Cox regression analysis to women with at least one previous smear. In the subsequent sensitivity analysis, we selected only women who attended the next screening round (within six years after a normal screening test result) in order to examine the effect that possible differences in the attendance rates at next screening might have had on the comparisons between the cytology tests. We repeated the latter analysis in the third sensitvity analysis with the addition of an extra confounding factor, the type of cytology test used in the subsequent screening round. We did this to correct for the potential differences in the sensitivity of the subsequent screening test.

Difference in CIN detection rates per 100,000 screening samples and 72-month cumulative cervical cancer incidence after 100,000 normal screening samples

We assessed the difference in CIN detection rates per 100,000 SurePath and 100,000 ThinPrep samples (compared with the CIN detection rates per 100,000 conventional cytology samples) and compared it with the difference in the 72-month cumulative cervical cancer incidence after 100,000 SurePath and ThinPrep normal screening samples. Information on the calculation of the difference in detection rates per 100,000 primary screening samples can be found in the **Appendix**. We calculated the 72-month cumulative cancer incidence rates for SurePath and ThinPrep by multiplying the distribution of the 72-month cumulative cancer incidence rate for conventional cytology with the distribution of the adjusted HRs for SurePath and ThinPrep versus conventional cytology, as obtained by Cox regression.

RESULTS

Within the follow-up period, 1042 invasive cervical cancers were diagnosed after 3,028,865 normal conventional cytology samples, 231 cancers were diagnosed after 1,303,817 normal SurePath samples, and 328 cancers were diagnosed after 1,591,792 normal ThinPrep samples (**Table 1**). This corresponds to 7.6, 4.8, and 6.3 cervical cancer diagnoses per 100,000 person years, respectively.

Crude cumulative incidence

Compared with conventional cytology, the 12-, 24-, 36-, 48-, 60- and 72-month cumulative incidences of invasive cervical cancer were significantly lower for SurePath (**Figure 1**). When we compared SurePath with ThinPrep, all but the 24-month cumulative incidences were significantly lower for SurePath. No significant differences were apparent between ThinPrep and conventional cytology. The 72-month cumulative incidence was 44.6 (95% CI: 37.8 to 52.6) after 100,000 normal SurePath samples, 58.5 (95% CI: 54.6 to 62.7) after 100,000 normal conventional cytology samples, and 66.8 (95% CI: 56.7 to 78.7) after 100,000 normal ThinPrep samples.

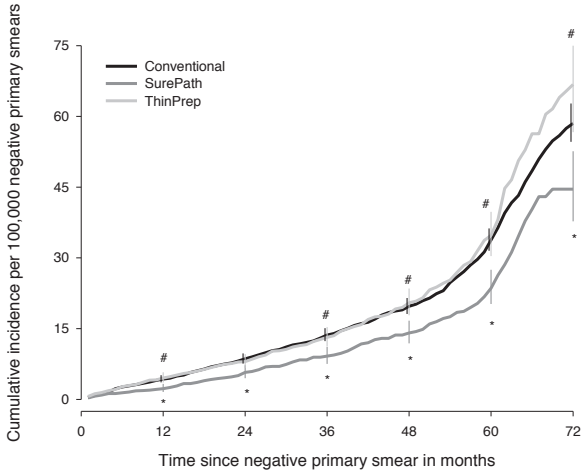


Figure 1. Cumulative cervical cancer incidence after 100,000 normal conventional cytology, SurePath, and ThinPrep samples taken within the Dutch cervical cancer screening program.

Cumulative cervical cancer incidence was calculated by Kaplan-Meier analyses. The 95% confidence intervals are depicted by vertical lines.

* Significant difference (P<0.05) between SurePath and conventional cytology.

\# Significant difference between SurePath and ThinPrep.

No significant differences between ThinPrep and conventional cytology were detected.

Distribution of potential confounding factors

The distribution of calendar time differed significantly between the methods of cytology testing (P<0.001). In 2000, 94% of the included normal screening samples consisted of conventional cytology, whereas by 2012 this percentage had dropped to 2% (**Figure 2**). We also observed a large significant difference for the distributions of follow-up time. For instance, almost 80% of the normal conventional cytology samples had a follow-up time of at least 48 months, whereas for SurePath and ThinPrep this was the case for slightly more than 50 and 35% of the normal samples, respectively. Small but significant differences were also present in the distributions of SES, screening history, and age (**Table 1**). Missing values were imputed for SES (1.4% of the primary normal samples had a missing value).

Table 1. Characteristics of normal screening samples and their follow-up for conventional cytology, SurePath, and ThinPrep. Values are numbers (percentages) unless stated otherwise.

	Conventional	SurePath	ThinPrep	P value
Normal screening samples	3,028,865	1,303,817	1,591,792	
Person years at risk	13,796,018	4,835,917	5,201,188	
Normal screening samples followed by subsequent screening[a]	1,931,397 (63.8)	445,726 (34.2)	370,519 (23.3)	<0.001
Median (interquartile range) normal screening samples per woman	1 (1-2)	1 (1-1)	1 (1-1)	<0.001
Invasive cervical cancers diagnosed after a normal screening sample	1,042	231	328	<0.001
Screen-detected[b]	414	84	103	<0.001
Clinically detected[c]	628	147	225	<0.001
Follow-up time:				<0.001
0-12 months	208,668 (6.9)	73,905 (5.7)	95,563 (6.0)	
12-24 months	105,945 (3.5)	191,027 (14.7)	321,784 (20.2)	
24-36 months	129,165 (4.3)	187,410 (14.4)	311,295 (19.6)	
36-48 months	203,768 (6.7)	189,063 (14.5)	284,262 (17.9)	
48-60 months	920,825 (30.4)	334,677 (25.7)	339,590 (21.3)	
60-72 months	1,460,494 (48.2)	327,735 (25.1)	239,298 (15.0)	
Age:				<0.001
29-33	411,873 (13.6)	167,015 (12.8)	193,998 (12.2)	
34-38	503,889 (16.6)	187,179 (14.4)	217,213 (13.6)	
39-43	516,728 (17.1)	218,559 (16.8)	267,194 (16.8)	
44-48	482,822 (15.9)	218,476 (16.8)	267,585 (16.8)	
49-53	434,620 (14.3)	192,594 (14.8)	240,801 (15.1)	
54-58	381,312 (12.6)	173,572 (13.3)	219,277 (13.8)	
59-63	297,621 (9.8)	146,422 (11.2)	185,724 (11.7)	
Screening history:				<0.001
No history[d]	396,174 (13.1)	167,880 (12.9)	194,251 (12.2)	
1 smear ≤7 yrs[e]	446,673 (14.7)	156,727 (12.0)	183,294 (11.5)	
1 smear >7 yrs[f]	35,164 (1.2)	15,388 (1.2)	20,003 (1.3)	
≥2 smears ≤7 yrs[g]	2,095,417 (69.2)	941,575 (72.2)	1,164,713 (73.2)	
≥2 smears >7 yrs[h]	55,437 (1.8)	22,247 (1.7)	29,531 (1.9)	
Socioeconomic status:				<0.001
Low	248,097 (8.2)	153,494 (11.8)	108,492 (6.8)	
Middle	2,501,696 (82.6)	1,038,602 (79.7)	1,337,521 (84.0)	
High	232,658 (7.7)	87,193 (6.7)	132,863 (8.3)	
Unknown	46,414 (1.5)	24,528 (1.9)	12,916 (0.8)	

[a] These differences are mainly caused by differences in follow-up time (see also **Figure 2**). Differences in length of follow-up were accounted for in all analyses. Sensitivity analyses were restricted to women with subsequent attendance at screening program.
[b] Include all cancers detected in first screening round following normal screening sample of ThinPrep, SurePath, or conventional cytology.
[c] Include all cancers detected outside screening program following normal screening sample of ThinPrep, SurePath, or conventional cytology.
[d] No history of cytological smears (inside or outside screening program) before normal screening sample.
[e] History of one cytological smear taken <7 years before normal screening sample.
[f] History of one cytological smear taken >7 years before normal screening sample.
[g] History of ≥2 cytological smears, last taken <7 years before normal screening sample.
[h] History of ≥2 cytological smears, last taken >7 years before normal screening sample.

Cox regression analyses of invasive cervical cancers

When we compared SurePath with conventional cytology, the hazard of an invasive cancer was significantly lower [HR of 0.81 (95% CI: 0.66 to 0.99)] (**Table 2**). This decreased hazard was mainly caused by a decreased hazard of a clinically detected cancer (i.e. not detected through program screening) [HR of 0.73 (95% CI: 0.57 to 0.93)]; the hazard of a screen-detected cancer was similar to that of conventional cytology [HR of 0.95 (95% CI: 0.72 to 1.27)].

When we compared ThinPrep with conventional cytology, the hazard of an invasive cancer was on average non-significantly higher [HR of 1.15 (95% CI: 0.95 to 1.38)]. This effect seemed to differ over time (*P*=0.063), with a HR of 0.95 (95% CI: 0.75 to 1.22) in the first 44

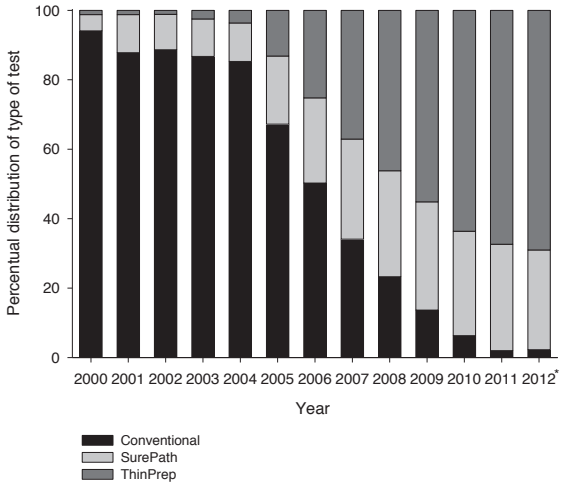


Figure 2. Annual distribution of the type of cytology used in normal screening samples taken within the Dutch cervical cancer screening program.[#]

* Until 31 March 2012.
[#] All normal screening samples taken within this time period (January 2000-March 2012) were included in the study, except if the type of cytology test was unknown.

months after the normal screening smear and of 1.40 (95% CI: 1.07 to 1.83) thereafter. This overall increased hazard was caused by an increased hazard of a screen-detected cancer [HR of 1.56 (95% CI: 1.17 to 2.08)], whereas the hazard of a clinically detected cancer was unaffected [HR of 0.96 (95% CI: 0.76 to 1.20)].

When we compared SurePath with ThinPrep, the hazard of an invasive cancer was significantly lower [HR of 0.71 (95% CI: 0.58 to 0.87)]. This decreased hazard was caused by both a decreased hazard of a clinically detected cancer [HR of 0.76 (95% CI: 0.59 to 0.97)] and a decreased hazard of a screen-detected cancer [HR of 0.61 (95% CI: 0.46 to 0.81)].

Sensitivity analyses

When selecting only women with at least one smear before the normal screening sample, we found that the HRs were consistent with the main analyses (not restricted to women with a screening history), although the effect of ThinPrep versus conventional cytology seemed to be somewhat less pronounced (**Table 2**).

When selecting only women who attended program screening within 72 months after a normal screening test result, we found that the hazard of a screen-detected cancer increased slightly for both SurePath and ThinPrep versus conventional cytology. The hazard of SurePath versus ThinPrep stayed similar as compared to the main analyses.

Of those women with conventional cytology at baseline followed by a subsequent screening round, 36% were re-screened with conventional cytology, 20% with SurePath, and 37% with ThinPrep. For the remaining 7%, the type of cytology test at re-screening was unknown. Of those with SurePath at baseline, 52% were re-screened with SurePath, 21% with ThinPrep, 4% with conventional cytology and 23% with an unknown type of test. Of those with ThinPrep at baseline, 55% were re-screened with ThinPrep, 6% with SurePath, 3% with conventional cytology, and 36% with an unknown type of test. The addition of the test method in the subsequent screening round as a confounding factor resulted in HRs similar to the ones in the second sensitivity analysis (without this extra confounder).

Difference in CIN detection rates and 72-month cumulative invasive cervical cancer incidence

The difference in the detection of CIN 2+ for SurePath and conventional cytology, as observed previously,[19] was consistent with the observed difference in cumulative incidence of cervical cancer after a normal screening sample. The use of SurePath as primary test method resulted in 94.4 (95% CI: 68.9 to 120.6) extra CIN diagnoses per 100,000 screening samples, whereas the 72-month cumulative incidence of cervical cancer decreased by 11.9 (95% CI: -15.6 to -4.2) (**Table 3**). The use of ThinPrep versus conventional cytology showed quite different results. Whereas the number of CIN diagnoses was similar to that with conventional cytology, the 72-month cumulative incidence of cervical cancer increased by 8.5 (95% CI: -0.7 to 18.8) after 100,000 normal screening test results.

Table 2. Hazard ratio of cervical cancer after a normal SurePath or ThinPrep screening sample as compared with after a normal conventional screening sample and as compared with each other, overall and stratified by the reason for the cervical examination that led to the cancer diagnosis.

	Main analyses		Sensitivity analyses: adjusted HR (95% CI)		
	Unadjusted HR (95% CI)	Adjusted HR (95% CI)	(1) Restricted to women with ≥1 previous smear	(2) Restricted to women attending program screening	(3) As for (2) + test method in subsequent screening round as confounding factor
All cancers					
SurePath vs. CC	0.74 (0.64 to 0.85)	0.81 (0.66 to 0.99)[a]	0.80 (0.64 to 0.99)	NA	NA
ThinPrep vs. CC	1.07 (0.95 to 1.22)	1.15 (0.95 to 1.38)[b]	1.09 (0.89 to 1.34)	NA	NA
SurePath vs. ThinPrep	0.69 (0.58 to 0.81)	0.71 (0.58 to 0.87)[c]	0.73 (0.58 to 0.91)	NA	NA
Clinically detected cancers					
SurePath vs. CC	0.70 (0.59 to 0.84)	0.73 (0.57 to 0.93)[d]	0.71 (0.54 to 0.92)	NA	NA
ThinPrep vs. CC	1.03 (0.88 to 1.20)	0.96 (0.76 to 1.20)[e]	0.91 (0.71 to 1.16)	NA	NA
SurePath vs. ThinPrep	0.68 (0.55 to 0.84)	0.76 (0.59 to 0.97)[f]	0.78 (0.60 to 1.02)	NA	NA
Screen-detected cancers					
SurePath vs. CC	0.80 (0.64 to 1.02)	0.95 (0.72 to 1.27)[g]	0.94 (0.68 to 1.29)	1.01 (0.72 to 1.41)	1.02 (0.71 to 1.47)
ThinPrep vs. CC	1.16 (0.93 to 1.44)	1.56 (1.17 to 2.08)[g]	1.58 (1.15 to 2.16)	1.65 (1.18 to 2.28)	1.62 (1.15 to 2.29)
SurePath vs. ThinPrep	0.70 (0.52 to 0.93)	0.61 (0.46 to 0.81)[g]	0.95 (0.43 to 0.81)	0.61 (0.44 to 0.86)	0.63 (0.43 to 0.90)

Differences in follow-up were taken into account and hazard ratios are shown unadjusted and adjusted for laboratory (included as random effect), age, socioeconomic status, screening history, and calendar time. Underlined = Significant. A P value of <0.05 was considered to be statistically significant.

HR = hazard ratio; CC = conventional cytology; NA = not applicable.

[a] HR was not time dependent; P = 0.559

[b] HR appeared to be time dependent, although not significantly; P = 0.063

[c] HR was not time dependent; P = 0.667

[d] HR was not time dependent; P = 0.658

[e] HR was not time dependent; P = 0.306

[f] HR was not time dependent; P = 0.954

[g] Time dependencies not calculated for screen-detected cancer, as they occur only approximately 60 months after baseline.

Table 3. CIN detection rates per 100,000 screening samples and 72-month cumulative cervical cancer incidence after 100,000 normal screening samples for conventional cytology, and the difference in those measures for SurePath and ThinPrep compared with conventional cytology.

	Conventional cytology	SurePath vs. conventional cytology	ThinPrep vs. conventional cytology
Number of CIN diagnoses per 100,000 screening samples			
CIN 1	216.1	+30.1 (+18.1 to +42.8)	-3.5 (-14.3 to +7.9)
CIN 2	220.0	+31.2 (+19.0 to +44.1)	+9.4 (-2.1 to +21.5)
CIN 3	495.0	+30.3 (+12.0 to +49.3)	-12.2 (-29.6 to +5.9)
Total CIN	931.0	+94.4 (+68.9 to +120.6)	-6.8 (-30.6 to +17.6)
Number of cervical cancer diagnoses after 100,000 normal screening samples[a]			
Total cancers	58.5 (54.6 to 62.7)	-11.9 (-15.6 to -4.2)	+8.5 (-0.7 to +18.8)

Numbers were corrected for confounding factors. The 95% confidence intervals are given in parentheses.
CIN = cervical intraepithelial neoplasia.

[a] Differences in the distribution of follow-up were taken into account, and laboratories were included as random effects in the model.

DISCUSSION

The risk of invasive cervical cancer was 19% lower after a normal SurePath sample than after a normal conventional cytological sample, which was mainly caused by a 27% lower risk of a clinically detected cancer (i.e. not detected through program screening). The use of SurePath resulted in 12 fewer cervical cancers per 100,000 normal screening samples, whereas the number of detected CIN lesions increased by 94. The risk of invasive cervical cancer seemed to be 15% higher for ThinPrep in comparison to conventional cytology, but the magnitude of the difference seemed to differ over time. Within the first 44 months after the normal screening sample, the risks were comparable; thereafter, the risk was 40% higher when using ThinPrep. Both the increased risk and the difference over time were mainly due to a 56% higher risk for a screen-detected cancer. The use of ThinPrep resulted in eight additional cervical cancers per 100,000 normal screening samples, whereas the number of detected CIN lesions was slightly but not statistically significantly lower.

Strengths and limitations

This study is the first to compare rates of invasive cervical cancer detected after a normal screening sample between two different types of LBC tests and conventional cytology, a widely accepted proxy for examining differences in the sensitivity to detect progressive CIN lesions. In addition, we examined the drawbacks of implementation of LBC by comparing indicators of overdiagnosis.

Our study has some limitations. First, we were not able to correct for the use of automated reading, although the possible influence would be small given that automated reading has been introduced in relatively few Dutch laboratories. Moreover, multiple studies have shown

that CIN 2+ detection was unaffected or slightly decreased by adding automated assisted reading to ThinPrep or SurePath.[33-35] Second, as we did not have a unique identification code [the identification code was based on the first eight letters of the (maiden) family name and birth date], tests belonging to different women may have been allocated to a single woman (so-called fusions). However, we think it unlikely that these fusions would be correlated with the type of cytology test used. Third, we did not have individual data on which type of primary test was used. Therefore, we used the date of the primary cytological smear and laboratory's conversion date fixed to the first date of the quarter to deduce which type of cytology test was used. This means that normal screening samples taken during this quarter may have been misclassified to some extent, leading to a slight underestimation of the effects. Fourth, we were not able to censor follow-up for death and migration. However, as both mortality and migration rates are relatively low at screening ages,[36,37] we do not expect that this has biased our results. Fifth, restricting our analyses to squamouscell carcinomas, adenocarcinomas, and/or micro-invasive and macro-invasive carcinomas was not possible, as this information in PALGA is not accurate and many values are missing. Finally, we did not correct for possible learning curve effects, as the aim of our study was to examine the effect of using SurePath and ThinPrep in routine practice, which also includes a possible learning effect.

Effect of confounding factors and sensitivity analyses

As only laboratory and calendar time were clearly correlated with the moment of implementation of LBC, their confounding effects were much more pronounced than those of age, screening history and SES. Although significantly different, their distribution differences were very small and, therefore, confounding effects were negligible. It is possible that we did not take into account the effect of other (unknown) potential confounders. The fact that no large differences in age, screening history, and SES is reassuring, however.

We found that the effect of ThinPrep compared with conventional cytology seemed to be somewhat less pronounced in women with a screening history compared with our main analysis (also including women without a screening history). This suggests that the risk of a cervical cancer after a normal Thinprep screening sample is perhaps increased more in women without versus with a screening history.

Restricting the analysis to women who attended the next screening round slightly increased the risk of a screen-detected cancer for both SurePath and ThinPrep, although the difference between these two types of LBC tests remained similar. This may have resulted in a slight underestimation of the HRs between SurePath and conventional cytology and between ThinPrep andconventional cytology.

In an ideal situation, the type of cytology test used would differ between the groups only at baseline to ensure that results are not biased by differences in sensitivity to detect a cervical cancer in the episode following the normal screening sample. Our finding that the addition of the second type of test (in the subsequent episode) as confounder did not change our results was reassuring.

Explanation of the main results

In our previous study, using the same data as in this study, we showed that the detection of CIN 2+ was increased by using SurePath compared with conventional cytology, whereas it was unaffected by using ThinPrep.[19] As the use of SurePath resulted in decreased rates of cervical cancer after a normal screening sample, this indicates that at least part of the extra detected CIN lesions were progressive. As the use of ThinPrep seemed to result in increased cancer rates, this suggests that fewer of the detected CIN lesions were progressive.

In addition, we showed that the use of SurePath was associated with lower rates ofcervical cancer after a normal screening sample compared with the use of ThinPrep, indicating that the sensitivity to detect progressive CIN lesions is higher for SurePath. These suggested differences in sensitivity are most likely caused by differences between the techniques of the LBC tests, such as the extent of fixation, the technique of taking a representative sample from the vial, and the retention of the brush (the collecting device) in the fluid.[38,39] Studies have shown that retaining the brush, as is done when using SurePath, is associated with an increased cell yield compared with rinsing and discarding the brush, as is done when using ThinPrep.[40,41]

Extrapolation of the results

The cumulative incidence of cancer after a normal sample seemed to be higher for ThinPrep than for conventional cytology mainly because the risk of a screen-detected cancer after a normal sample was higher. In general, screen-detected cancers are found at a lower stage than clinically diagnosed ones, so their survival is probably better.[42] Therefore, the suggested negative effect of using ThinPrep is probably less pronounced for cervical cancer mortality than it is for incidence. The opposite is true for the positive effect of Surepath, as we found that SurePath was primarily protective for clinically detected cancers. As no data on mortality were available, we were not able to estimate the effects of implementation of LBC on cervical cancer mortality. Although our results may not seem very relevant for the future of the Dutch cervical cancer screening program, as primary cytology screening has recently been replaced by primary HPV screening, they certainly can be relevant to other countries with organized primary cytology screening programs that have switched to using SurePath and/or ThinPrep or will switch in the near future.

Overdiagnosis

An important drawback of cervical cancer screening is the overdiagnosis and overtreatment of CIN lesions (i.e. of lesions that would never have progressed to clinical cervical cancer in the absence of screening). The possible increase in such overdiagnosis and overtreatment related to more sensitive screening should be taken into account when considering new screening options. With SurePath, the prevention of 12 extra cervical cancers within the first six years after screening was accompanied by the detection of 94 extra CIN lesions at that screening round. Most of these CIN lesions would never have become invasive cancer and could therefore be classified as overdiagnosis. Assuming that only CIN 2+ are treated, replac-

ing conventional cytology with Surepath would have implied that roughly five more CIN treatments are performed to prevent one additional cervical cancer diagnosis. However, in a subsequent screening round for the same cohort, increased detection rates tend to wane, and so does overdiagnosis and overtreatment. This effect was also observed in randomized controlled trials in which cytology screening was replaced by highly sensitive HPV-based screening in the intervention arm. The detection of high-grade CIN was increased at the prevalence round but decreased at the subsequent incidence round, accompanied by a reduced risk of interval cancers.[43,44] Therefore, in our study, the number of additional CIN treatments per additionally prevented cancer is, if anything, overestimated.

Comparison with the literature

Whether one cytology test is preferred over another should depend not only on the sensitivity to detect progressive CIN lesions and rates of overdiagnosis but also on factors such as the possibility to test for the presence of HPV in the residual material and the percentage of unsatisfactory smears. Fontaine et al. have shown that unsatisfactory rates are significantly lower when using SurePath instead of ThinPep.[45] Although this was not shown in our previous study,[19] we then found similar results (an odds ratio of 0.74 (95% CI: 0.72 to 0.75)) when comparing unsatisfactory rates between SurePath and ThinPrep.

Conclusion

The six-year cumulative incidence of cervical cancer after a normal screening sample was significantly lower for Surepath than for conventional cytology and ThinPrep, strongly suggesting that the sensitivity of Surepath to detect progressive CIN lesions is higher. The use of ThinPrep compared with the use of conventional cytology seemed to be associated with a higher cumulative cancer incidence, suggesting that the sensitivity to detect progressive CIN lesions is lower, although results were non-significant. Our findings should provoke reconsideration of the assumed similarity in the sensitivity for progressive CIN between the different types of LBC tests and conventional cytology.

REFERENCES

1. Rask J, Lynge E, Franzmann M, et al. Impact of technology on cytology outcome in cervical cancer screening of young and older women. Int J Cancer 2014;134:2168-79.
2. Public Health England. Cervical screening: A pocket guide. Sheffield: NHS Cervical Cancer Screening Programme, 2009.
3. Cox JT. History of the use of HPV testing in cervical screening and in the management of abnormal cervical screening results. J Clin Virol 2009;45 Suppl 1:S3-S12.
4. Albrow R, Kitchener H, Gupta N, et al. Cervical screening in England: the past, present, and future. Cancer Cytopathol 2012;120:87-96.
5. Colgan TJ, McLachlin CM, Cotterchio M, et al. Results of the implementation of liquid-based cytology-SurePath in the Ontario screening program. Cancer 2004;102:362-7.
6. Fremont-Smith M, Marino J, Griffin B, et al. Comparison of the SurePath liquid-based Papanicolaou smear with the conventional Papanicolaou smear in a multisite direct-to-vial study. Cancer 2004;102:269-79.
7. Beerman H, van Dorst EB, Kuenen-Boumeester V, et al. Superior performance of liquid-based versus conventional cytology in a population-based cervical cancer screening program. Gynecol Oncol 2009;112:572-6.
8. Siebers AG, Klinkhamer PJ, Arbyn M, et al. Cytologic detection of cervical abnormalities using liquid-based compared with conventional cytology: a randomized controlled trial. Obstet Gynecol 2008;112:1327-34.
9. Sigurdsson K. Is a liquid-based cytology more sensitive than a conventional Pap smear? Cytopathology 2013;24:254-63.
10. Arbyn M, Bergeron C, Klinkhamer P, et al. Liquid compared with conventional cervical cytology: a systematic review and meta-analysis. Obstet Gynecol 2008;111:167-77.
11. Whitlock EP, Vesco KK, Eder M, et al. Liquid-based cytology and human papillomavirus testing to screen for cervical cancer: a systematic review for the U.S. Preventive Services Task Force. Ann Intern Med 2011;155:687-97, W214-5.
12. Coste J, Cochand-Priollet B, de Cremoux P, et al. Cross sectional study of conventional cervical smear, monolayer cytology, and human papillomavirus DNA testing for cervical cancer screening. BMJ 2003;326:733.
13. Hussein T, Desai M, Tomlinson A, et al. The comparative diagnostic accuracy of conventional and liquid-based cytology in a colposcopic setting. BJOG 2005;112:1542-6.
14. Ronco G, Cuzick J, Pierotti P, et al. Accuracy of liquid based versus conventional cytology: overall results of new technologies for cervical cancer screening: randomised controlled trial. BMJ 2007;335:28.
15. Siebers AG, Klinkhamer PJ, Grefte JM, et al. Comparison of liquid-based cytology with conventional cytology for detection of cervical cancer precursors: a randomized controlled trial. JAMA 2009;302:1757-64.
16. Taylor S, Kuhn L, Dupree W, et al. Direct comparison of liquid-based and conventional cytology in a South African screening trial. Int J Cancer 2006;118:957-62.
17. Ferenczy A, Robitaille J, Franco E, et al. Conventional cervical cytologic smears vs. ThinPrep smears. A paired comparison study on cervical cytology. Acta Cytol 1996;40:1136-42.
18. Sykes PH, Harker DY, Miller A, et al. A randomised comparison of SurePath liquid-based cytology and conventional smear cytology in a colposcopy clinic setting. BJOG 2008;115:1375-81.

19. Rozemeijer K, Penning C, Siebers AG, et al. Comparing SurePath, ThinPrep, and conventional cytology as primary test method: SurePath is associated with increased CIN II+ detection rates. Cancer Causes Control 2016;27:15-25.
20. Rebolj M, Rask J, van Ballegooijen M, et al. Cervical histology after routine ThinPrep or SurePath liquid-based cytology and computer-assisted reading in Denmark. Br J Cancer 2015;113:1259-74.
21. Rebolj M, van Ballegooijen M, van Kemenade F, et al. No increased risk for cervical cancer after a broader definition of a negative Pap smear. Int J Cancer 2008;123:2632-5.
22. Casparie M, Tiebosch AT, Burger G, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol 2007;29:19-24.
23. Rebolj M. Recent developments in the Dutch cervical cancer screening programme. Public Health. Rotterdam: Erasmus University Rotterdam, 2008.
24. Blok M. [Description of the Paleba System and its Algorithms 2012]. In: Palga, ed, 2013.
25. Dutch Society for Pathology. [Practice Guideline version 3.2 - For quality assurance of cytopathological research of the cervix], 2012.
26. The Netherlands Institute for Social Research. [From high to low; from low to high - The social-spatial development of quarters between 1971 and 1995]. Den Haag, 1998.
27. Vesco KK, Whitlock EP, Eder M, et al. Risk factors and other epidemiologic considerations for cervical cancer screening: a narrative review for the U.S. Preventive Services Task Force. Ann Intern Med 2011;155:698-705, W216.
28. Rozemeijer K, van Kemenade FJ, Penning C, et al. Exploring the trend of increased cervical intraepithelial neoplasia detection rates in the Netherlands. J Med Screen 2015.
29. Froment MA, Gomez SL, Roux A, et al. Impact of socioeconomic status and ethnic enclave on cervical cancer incidence among Hispanics and Asians in California. Gynecol Oncol 2014;133:409-15.
30. Jensen KE, Hannibal CG, Nielsen A, et al. Social inequality and incidence of and survival from cancer of the female genital organs in a population-based study in Denmark, 1994-2003. Eur J Cancer 2008;44:2003-17.
31. Harrell F. Regression Modeling Strategies - With Applications to Linear Models, Logistic Regression, and Survival Analysis: Springer-Verlag New York, 2001.
32. Therneau TM. coxme: Mixed Effects Cox Models, 2015.
33. Klug SJ, Neis KJ, Harlfinger W, et al. A randomized trial comparing conventional cytology to liquid-based cytology and computer assistance. Int J Cancer 2013;132:2849-57.
34. Palmer TJ, Nicoll SM, McKean ME, et al. Prospective parallel randomized trial of the MultiCyte ThinPrep((R)) imaging system: the Scottish experience. Cytopathology 2013;24:235-45.
35. Kitchener HC, Blanks R, Dunn G, et al. Automation-assisted versus manual reading of cervical cytology (MAVARIC): a randomised controlled trial. Lancet Oncol 2011;12:56-64.
36. [Mortality: What is the current situation?]. Volume 2016: National Institute of Public Health and the Environment.
37. [Migration: What is the current situation?]. Volume 2016: National Institute of Public Health and the Environment.
38. Zhao FH, Hu SY, Bian JJ, et al. Comparison of ThinPrep and SurePath liquid-based cytology and subsequent human papillomavirus DNA testing in China. Cancer Cytopathol 2011;119:387-94.
39. Thiryayi SA, Marshall J, Rana DN. An audit of liquid-based cervical cytology screening samples (ThinPrep and SurePath) reported as glandular neoplasia. Cytopathology 2010;21:223-8.
40. Bigras G, Rieder MA, Lambercy JM, et al. Keeping collecting device in liquid medium is mandatory to ensure optimized liquid-based cervical cytologic sampling. J Low Genit Tract Dis 2003;7:168-74.

41. Umana A, Dunsmore H, Herbert A, et al. Are significant numbers of abnormal cells lost on the discarded ThinPrep(R) broom when used for cervical cytology? Cytopathology 2013;24:228-34.
42. van der Aa MA, Schutter EM, Looijen-Salamon M, et al. Differences in screening history, tumour characteristics and survival between women with screen-detected versus not screen-detected cervical cancer in the east of The Netherlands, 1992-2001. Eur J Obstet Gynecol Reprod Biol 2008;139:204-9.
43. Arbyn M, Ronco G, Anttila A, et al. Evidence regarding human papillomavirus testing in secondary prevention of cervical cancer. Vaccine 2012;30 Suppl 5:F88-99.
44. Ronco G, Dillner J, Elfstrom KM, et al. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. Lancet 2014;383:524-32.
45. Fontaine D, Narine N, Naugler C. Unsatisfactory rates vary between cervical cytology samples prepared using ThinPrep and SurePath platforms: a review and meta-analysis. BMJ Open 2012;2:e000847.

APPENDIX: CALCULATION OF DIFFERENCES IN DETECTION RATES

In this appendix, we describe the methods that were used to determine the difference in CIN detection rates per 100,000 primary screening samples.

Selecting data from PALGA: CIN lesions

We identified primary samples taken within the national cervical cancer screening program between January 2000 and December 2011. As data until March 2013 were available to us, a minimum duration of 15 months follow-up was ensured. Histologically confirmed CIN lesions were identified by selecting all PALGA records that included corresponding pathology codes. Subsequently, lesions were linked to the type of cytology test used. Age, screening region, SES, and calendar year at the time of the primary sample were assessed in similar ways as in the main analysis.

Statistical analyses: CIN lesions

We compared CIN detection rates per 100,000 SurePath and 100,000 ThinPrep samples with CIN detection rates per 100,000 conventional cytology samples. As confounding factors are present, comparing observed CIN detection rates was not sufficient. Therefore, we calculated CIN detection rates per 100,000 SurePath and ThinPrep samples by multiplying the observed CIN detection rates per 100,000 conventional cytology samples with the adjusted odds ratios for SurePath and ThinPrep versus conventional cytology, as obtained in our previous study (**Table 1**).[1] These odds ratios were adjusted for differences in the distribution of age, screening region, SES, and calendar time between the three cytology tests.

Table 1. Factors to calculate the adjusted CIN detection rates for SurePath and ThinPrep. Given factors are odds ratios comparing SurePath and ThinPrep with conventional cytology, adjusted for age, screening region, SES, and calendar time.

	SurePath vs. CC (95% CI)	ThinPrep vs. CC (95% CI)
CIN 1	1.14 (1.08 to 1.20)	0.98 (0.93 to 1.04)
CIN 2	1.14 (1.09 to 1.20)	1.04 (0.99 to 1.10)
CIN 3	1.06 (1.02 to 1.10)	0.98 (0.94 to 1.01)
Total CIN	1.10 (1.07 to 1.13)	0.99 (0.97 to 1.02)

CC = conventional cytology; CIN = cervical intraepithelial neoplasia.

References

1. Rozemeijer K, Penning C, Siebers AG, Naber SK, Matthijsse SM, van Ballegooijen M, et al. Comparing SurePath, ThinPrep, and conventional cytology as primary test method: SurePath is associated with increased CIN II detection rates. Cancer Causes Control 2015.

Chapter 6

Offering self-sampling to non-attendees of organized primary HPV screening: when do harms outweigh the benefits?

Kirsten Rozemeijer, Inge M.C.M. de Kok, Steffie K. Naber, Folkert J. van Kemenade, Corine Penning, Joost van Rosmalen, Marjolein van Ballegooijen

Cancer Epidemiol Biomarkers Prev. 2015 Apr; 24(5); 1–10.

ABSTRACT

Background: Human papillomavirus (HPV) self-sampling might be a promising tool to increase effectiveness of primary HPV screening programs when offered to non-attendees. However, effectiveness could decrease if regular attendees "switch" to self-sampling, because self-sampling test characteristics may be inferior. We examined under which conditions the harms would outweigh the benefits.

Methods: The MISCAN-Cervix model was used to estimate quality-adjusted life years (QALY) gained and costs of offering HPV self-sampling to non-attendees. We varied the relative CIN 2+ sensitivity and specificity (self-sampling vs. regular sampling), extra attendance, risk of extra attendees, and the switching percentage.

Results: Without switching, offering self-sampling is (cost-)effective under every studied condition. If the attendance due to self-sampling increases by ≥6 percentage points, higher primary background risk women (unscreened women who will never attend regular screening) attend and the relative CIN 2+ sensitivity and specificity are ≥0.95; it is (cost-)effective to offer self-sampling to non-attendees, even if all regular attendees switch. If the relative sensitivity decreases to 0.90 combined with either a 3 percentage points extra attendance or the absence of higher primary background risk women, QALYs are lost when more than 30% to 20% of the regular attendees switch.

Conclusions: Offering self-sampling will gain health effects if the relative CIN 2+ sensitivity is ≥0.95, unscreened attendees are recruited, and the total attendance increases by ≥6 percentage points. Otherwise, switching of regular attendees may decrease the total effectiveness of the program.

Impact: Self-sampling needs to be implemented with great care and advantages of office-based sampling need to be emphasized to prevent switching.

INTRODUCTION

In the Netherlands, cervical cancer incidence and mortality have decreased in the past decades to 6.5 and 1.3 per 100,000 woman years (age-adjusted to the World Population) in 2012.[1] The introduction and improvements of the screening program played a considerable role in this decrease,[2] Since 1996, Dutch women of ages 30 to 60 years are invited to attend cervical cancer screening every five years. From 2016 onward, primary cytology will be replaced by primary high-risk human papillomavirus (HPV) testing,[3] because the sensitivity for detecting CIN 2+ lesions is higher when using HPV testing[4] and HPV testing can be performed on self-samples.[5,6] Although the current screening participation rate ranges from 65% to almost 70%,[7] it has been estimated that more than half of the invasive cervical cancers occur in women who did not participate in the previous six years. Moreover, some of these women had never been screened at all.[8] This shows that addressing non-attendance can increase the effectiveness of the program considerably.

Self-sampling devices, with which women can collect cervical cells themselves, have been developed recently. As self-sampling is more woman-friendly and less time consuming than letting a clinician, general practitioner, or midwife collect cervical cells, it probably increases participation in screening. Indeed, the Dutch PROHTECT study has shown that offering a self-sampling HPV test to non-attendees of the program increased the overall screening participation rate by about 6 percentage points.[9,10] However, the gain in effectiveness of the program [i.e., gain in quality-adjusted life years (QALYs)] probably not only depends on the increase in attendance, but also on the test characteristics of HPV self-sampling and on the ability to target higher risk non-attendees. It is likely that *unscreened women* (who were invited at least once but were never screened) have higher risks on developing cervical cancer than *one-time non-attendees* (who missed the last screening round, but have been screened in the past). Nevertheless, including any non-attendee will probably increase the effectiveness of the program. However, "switching" of *regular attendees* from office-based to self-sampling could, given a loss in detection (i.e., more loss to follow-up, possible lower sensitivity), result in a decrease of the effectiveness of the program (i.e., losing QALYs). In other words, the QALYs gained by attracting non-attendees could be annulled by the QALYs lost by switching of regular attendees. It is unclear at which level of switching this will happen.

The aim of this study is to examine the effectiveness of offering HPV self-sampling to non-attendees of a primary HPV screening program. We modeled effects of parameters such as the relative CIN 2+ sensitivity and specificity (self-sampling vs. regular sampling), the extra attendance via self-sampling, and the risk of extra attendees. Given that the percentage of women who will switch from office-based to self-sampling is unknown, we determined the percentage of switching that would result in a decrease of the total effectiveness of the program (i.e., harms outweigh the benefits, QALYs are lost). We also examined the circumstances (i.e., limits) under which it would not be cost-effective to offer HPV self-sampling to non-attendees.

METHODS

We used the MISCAN-Cervix model to estimate benefits, harms, and costs of offering a self-sampling HPV test to non-attendees.[11] For detailed information on the model specifications, see the **Model appendix**.

Assumptions for screening and triage

The screening policy considered is primary HPV screening with cytology triage, as will be implemented in the Netherlands (**Figure 1**).[12] Women will be invited for screening at ages 30, 35, 40, 50, and 60 years. In addition, women will be invited at ages 45, 55, and 65 years if they did not attend screening or had a positive HPV test in the previous screening round.

Assumptions for attendance

For HPV office-based sampling, we assumed an age-dependent overall 65% attendance rate as currently observed within the Dutch cytological screening program.[7] On the basis of the findings of the PROHTECT trial (offering self-sampling to non-attendees after an opting-out letter), we assumed that a self-sampling kit was sent to 85% of the non-attendees,[13] which resulted in an extra overall attendance of 6 percentage points.[9] We assumed that 29% of these extra-attendees are higher primary risk women (i.e., unscreened women who will never attend via office-based sampling and who have a 1.7 times higher primary background risk for developing cervical cancer than women who are willing to attend office-based sampling), which is equal to the proportion in non-attendees (i.e., 10% / 35% = 29%) (**Figure 2**). In addition to their increased primary background risk, these women also have an increased cervical cancer risk due to never attending regular screening.

We assumed that the loss to follow-up after a positive self-sampling test was higher than after a positive office-based sampling test. On the basis of the observed data, we assumed that 92% of the women comply with the first triage invitation and 68% with the second.7 With office-based sampling, the collected material can be used both for primary HPV and direct cytology triage testing (co-collection). Therefore, the first and only triage invitation is six months after the positive screening test. This results in a compliance of 100% for immediate cytology triage testing and 92% for triage testing six months after the positive office-based sampling test. In case of self-sampling, co-collection is not possible and women receive their first and second triage invitation directly and six months after the positive screening test. This results in a compliance of 92% for immediate cytology triage testing and 68% for triage testing six months after the positive self-sampling test (**Figure 1**).

As no data were available, we considered the two most extreme "switching" scenarios in the base-case analyses: no regular attendees and all regular attendees switch from office-based to self-sampling.

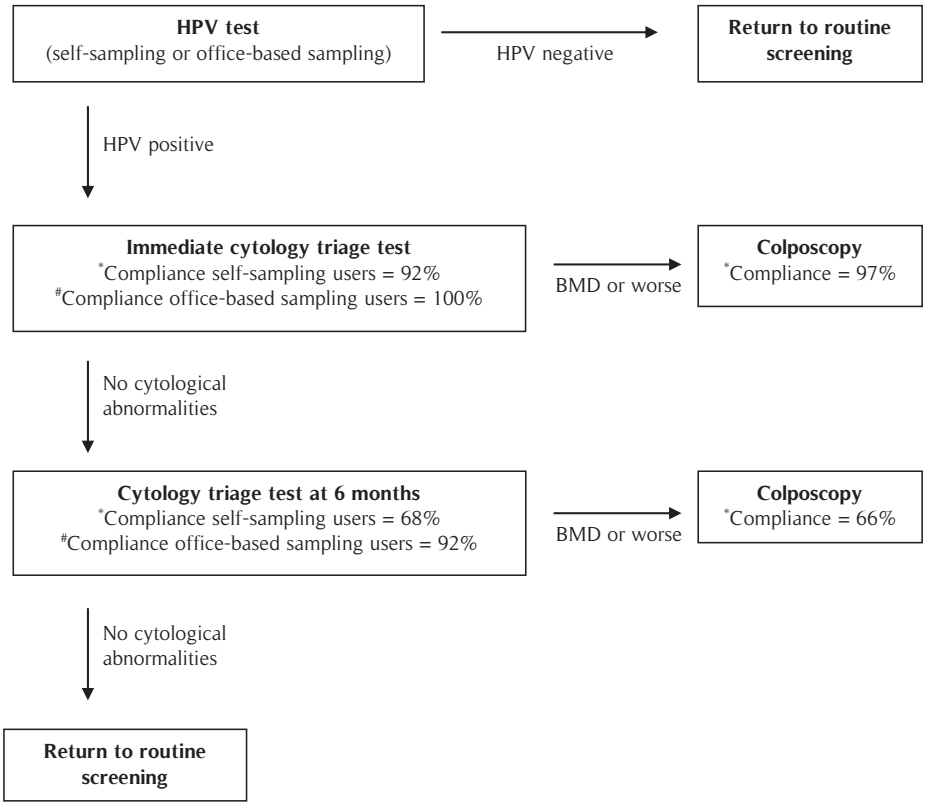


Figure 1. Triage strategy and compliance assumptions after a positive self-sampling and office-based sampling HPV test. We assumed that the compliance (i.e., attendance for triage and colposcopy) behavior does not differ between self-sampling, future, and current office-based sampling users.

* Compliance rates of the first triage test (i.e., immediate cytology triage test), second triage test (i.e., cytology triage test at 6 months), and of colposcopy are assumed equal to those observed within the current program.

\# First triage (i.e., immediate cytology triage) after a positive office-based sampling test will be performed using co-collection, so the compliance will automatically be 100%. Compliance with the second triage test (i.e., cytology triage at 6 months) is assumed to be equal to the first triage test in the current program.

HPV = human papillomavirus; BMD = borderline or mildly dyskaryosis.

Base-case assumptions for test characteristics

The test characteristics of self-sampling were based on the assumption that a validated PCR test was used, as for instance the GP5+/6+.[4,14] According to the recent meta-analysis of Arbyn and colleagues,[15] the point estimate for the relative sensitivity of CIN 2+ when comparing self-sampling with office-based sampling is approximately 0.95, whereas the point estimate for the relative specificity is probably higher than 1.00. Therefore, we assumed a 5 percentage points lower sensitivity for high-risk HPV infections when self-sampling (i.e. 80 vs. 85%), and an equal specificity of 100% (i.e. the true but uncertain value of specificity is probably somewhat lower than 100% due to cross-reactivity with low-risk HPV types and

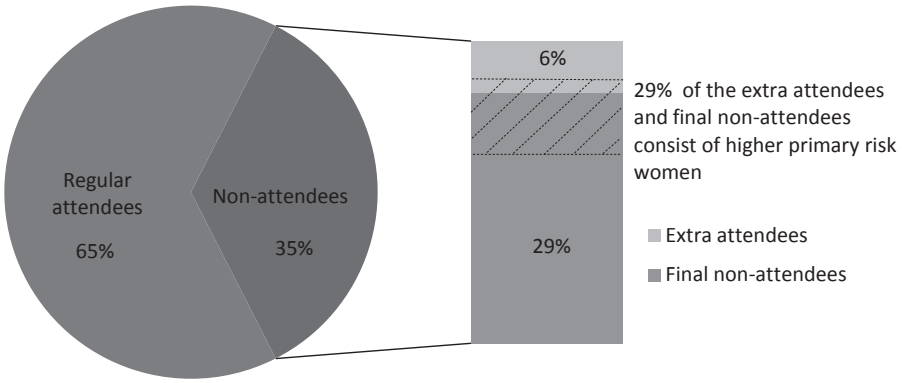


Figure 2. Distribution of regular attendees, non-attendees, and extra attendees within the screening population.

After receiving a screening invitation 65% of the invited women will attend via office-based sampling (i.e., regular attendees) and 35% will not attend (i.e., non-attendees). After an HPV self-sampling test has been offered to the non-attendees, 17% of them will attend (i.e., extra attendees; = 6% of the screening population) and 83% will not (i.e., final non-attendees; = 29% of the screening population). 29% of the non-attendees consist of higher primary risk women (= 10% of the screening population). We assumed that the proportion of higher primary risk women in the extra attendees and final non-attendees stayed equal to that in the non-attendees (= 1.7% of the screening population are higher primary risk women who attend via self-sampling, 8.3% are higher primary risk women who do not attend). Office-based sampling users consist of regular attendees, while self-sampling users consist of extra attendees and, in case of switching, of (part of the) regular attendees.

contamination). By including fast clearing high-risk HPV infections, we were able to model a lack of specificity.

As women in our model can have multiple lesions at the same time, the CIN 2+ sensitivity not only depends on the sensitivity for a high-risk HPV infection, but also on the specificity. Therefore, a 5 percentage points lower sensitivity for high-risk HPV infections and an equal specificity corresponds with a 0.95 relative CIN 2+ sensitivity. On the other hand, the specificity for a CIN 2+ lesion depends on the specificity and sensitivity for a high-risk HPV infection. As the prevalence of high-risk HPV infections in women without CIN 2+ is higher in young women and relatively more young women use self-sampling, a 5 percentage points lower sensitivity for high-risk HPV infections and an equal specificity corresponds with a 0.99 relative CIN 2+ specificity.

Assumptions for costs and utilities

Table 1 presents the inputs for utilities and costs used in the analyses. Utilities were based on (inter)nationally published data.[16] The unit costs were estimated from a societal perspective. As compared to office-based sampling, self-sampling was assumed to be less expensive, but the costs of immediate cytology triage were higher. Diagnostic costs of women referred for colposcopy, treatment costs and costs of palliative care were equal between the two tests and were derived from previous cost studies performed in the Netherlands.[17]

Cost-effectiveness analysis

We assumed that the evaluated alternative screening policies (i.e. primary HPV screening with and without offering HPV self-sampling to non-attendees) started in 2013 and continued until all women reached the final screening age. The costs and effects of the simulated screening programs were counted from 2013 onward until all simulated women (i.e. born between 1953 and 1992) had died. We also simulated the last three screening rounds before 2013 (i.e. primary cytology screening with cytology triage), because they can influence the effectiveness of the screening program after 2013. We simulated ten million women for each strategy. Future costs and health effects [life years (LYs) lived and utility losses] were discounted towards the year 2013 at an annual rate of 3%. We computed the net costs and number of QALYs gained by screening as the differences between the simulations with and without screening. The incremental cost-effectiveness ratio (ICER) was defined as the increase in costs per additional (QA)LY gained when self-sampling would be offered to non-attendees as compared to no such offer. The cost-effectiveness threshold was set to €20,000 per QALY gained, based on decisions of the Dutch government,[18] and to €50,000, which is often used in an international perspective.[19]

Multivariate sensitivity analyses

The relative CIN 2+ sensitivity and specificity can differ from the estimates we used in our base-case analysis, as there is uncertainty about the true value [e.g. the 95% confidence interval (CI) for the pooled relative sensitivity and specificity when using the GP5+/6+ is 0.89 to 1.01 and 0.95 to 1.29 respectively].[15] In addition, they depend on the type of HPV DNA test used,[15] meaning that the values could be different when another validated HPV DNA test is used. Therefore, we choose to set the sensitivity for a high-risk HPV infection equally, 5, and 10 percentage points lower for self-sampling as compared to office-based sampling. The specificity was set equally, 5, and 15 percentage points lower.

As the CIN 2+ sensitivity and specificity depend on both the sensitivity and specificity for high-risk HPV infections, the CIN 2+ sensitivity and specificity varied slightly between different combinations of self-sampling test characteristics for high-risk HPV infections. This resulted in a relative CIN 2+ sensitivity that varied between 0.89 and 1.02, and a relative CIN 2+ specificity that varied between 0.84 and 1.00.

The relative CIN 2+ sensitivity and specificity are expected to have a major influence on the effectiveness of the program, especially when women switch. Therefore, we determined the percentage of women switching (0%, 10%, …, 90%, 100%) for which offering self-sampling is no longer effective (i.e. QALYs are lost) or cost-effective (i.e. ICER is larger than the cost-effectiveness threshold).

In addition, we varied the loss in quality of life associated with cytology triage, the costs of the self-sampling kit, the extra attendance via self-sampling, and the attendance of higher primary risk women women (i.e. unscreened women who will never attend office-based sampling) and their background risk for cervical cancer.

Table 1. Model input: Costs and utilities under base-case assumptions.

		Utility loss	
Parameter	**Costs in €**	**Fraction**	**Duration**
Invitation	4.85		
Primary office-based sampling test		0.006	2 weeks
Program[a]	2.68 / 2.95		
Organisation	12.50		
Office-based sampling	12.09		
Laboratory	29.00		
Time/travel	6.28		
Total	62.55 / 62.82		
Primary self-sampling test		0.006	2 weeks
Self-sampling kit[b]	6.00		
Program	2.68		
Organisation	12.50		
Laboratory	29.00		
Time[c]	2.76		
Total	50.94		
Immediate cytology triage test after positive office-based sampling[d]		NA	NA
Laboratory[e]	30.27		
Total	30.27		
Immediate cytology triage test after positive self-sampling		0.006	2 weeks
Organisation	10.00		
Office-based sampling	12.09		
Laboratory	32.27		
Time/travel	6.28		
Total	60.64		
Cytology triage test at 6 months		0.006	0.5 year
Organisation	10.00		
Office-based sampling	12.09		
Laboratory	32.27		
Time/travel	6.28		
Total	60.64		
Diagnosis and treatment of pre-invasive stages			
False-positive referral	296	0.005	0.5 year
CIN grade 1	924	0.03	0.5 year
CIN grade 2	1,368	0.07	1 year
CIN grade 3	1,602	0.07	1 year

Table 1 (continued). Model input: Costs and utilities under base-case assumptions.

		Utility loss	
Parameter	**Costs in €**	**Fraction**	**Duration**
Diagnosis and treatment of cancer			
FIGO 1A	5,246	0.062	5 years
FIGO 1B	12,440	0.062	5 years
FIGO 2+ (screen detected)	12,261	0.28	5 years
FIGO 2+ (clinically detected)	11,451	0.28	5 years
Terminal care	27,859	0.712	1 month

Costs are in 2012 prices. NA = not applicable.

[a] As the total program costs were fixed, the costs per test was dependent on the number of women participating in the screening program. As this number was higher with the inclusion of the self-sampling test, the costs per test were lower in the situation with versus without self-sampling.

[b] We assumed that 85% of the non-attendees received the self-sampling kit of €6.00 at home, irrespective of whether they used it or not. This price was estimated based on personal communication with multiple developers of brush and lavage HPV self-sampling kits. The remaining costs (e.g. laboratory, organisation, etc.) were only taking into account among women who actually attended via self-sampling.

[c] Given that it was not required to go to the general practitioner's office, we assumed that women who attended via self-sampling spent half of the time to screening (€2.76 instead of €5.52) as compared to women who attended via office-based sampling, while travel costs (€0.76) were absent.

[d] Co-collection based analysis was possible after positive office-based sampling and, therefore, women did not have to go to the general practitioner´s office for the immediate cytology triage test.

[e] We assumed that part of the material costs (€2.00) were already included in the price of the office-based sampling test. Therefore, laboratory costs of immediate cytology triage after a positive office-based sampling test were lower than after a positive self-sampling test.

Utility loss associated with cytology triage. True estimates of the utility loss due to having cytology triage are unavailable. Especially if self-sampling is associated with a lower specificity, this may influence the effectiveness of offering self-sampling. Therefore, we studied the effect of assuming no utility loss to 0.012 per week for being in triage (base case: 0.006 per week).

Costs. The total price of a self-sampling kit depends on many factors (e.g. type of self-sampling device, possibility to achieve economies of scale, and on-going innovations for the self-sampling test). Therefore, we varied unit self-sampling kit costs from €3.50 to €10.00 (base case: €6.00).

Attendance via self-sampling. We varied the extra attendance rate due to self-sampling from 3 to 10 percentage points (base case: 6 percentage points). Furthermore, we varied the proportion of higher primary risk women in extra attendees from 0 to 50% (base case: 29%).

Background risk for cervical cancer of "higher primary risk" women. We assumed that all women have the same background risk for cervical cancer (base case: "higher primary risk" women have a 1.7 times higher background risk than regular attendees), although "higher primary risk" women still have an increased cervical cancer risk due to never attending regular screening.

RESULTS

Base-case scenario

Table 2 presents the undiscounted effects and costs per 100,000 simulated women when offering HPV self-sampling to non-attendees of a primary HPV screening program. Without switching, offering self-sampling increased the number of triage tests and false-positive referrals for colposcopy (+7.5% and +5.7%, respectively) and decreased the number of cervical cancer cases and deaths by 7.0% and 9.2%, respectively. Because the costs increased by only 5.5%, it was not only effective (+12.1% QALYs gained) but also cost-effective (ICER of €2,115 per QALY gained) to add self-sampling to the program (**Table 3**).

As the sensitivity of self-sampling was lower than that of office-based sampling and because the probability of being lost to follow-up after a positive self-sampling test was higher than after a positive office-based sampling test, switching resulted in a decrease of the number of triage tests and subsequently false-positive referrals, an increase of the number of cervical cancers, and a decrease in the number of cervical cancer deaths prevented

Table 2. Undiscounted simulated effects and costs, compared to no screening, of primary HPV screening with and without offering self-sampling to non-attendees under base-case assumptions, per 100,000 simulated women.

	Without self-sampling	With self-sampling (difference with situation without self-sampling, in %)			
		No switching		100% switching	
Effects, No.					
Primary screens	219,953	234,171	(+6.5)	234,108	(+6.4)
Triage tests	12,983	13,952	(+7.5)	10,834	(-16.6)
False-positive referrals	163	173	(+5.7)	133	(-18.5)
CIN grade 1 diagnoses	786	842	(+7.2)	676	(-13.9)
CIN grade 2 diagnoses	523	565	(+8.1)	460	(-12.0)
CIN grade 3 diagnoses	844	924	(+9.5)	812	(-3.7)
CeCa cases	626	582	(-7.0)	631	(+0.8)
Screen-detected CeCa cases	77	86	(+12.2)	89	(+16.0)
Clinically-detected CeCa cases	549	496	(-9.6)	541	(-1.3)
CeCa deaths	250	227	(-9.2)	247	(-1.3)
LYs lost	5929	5388	(-9.1)	5833	(-0.8)
QALYs lost	772	735	(-4.8)	760	(-1.6)
Costs, €					
Testing costs	16,022,798	17,134,705	(+6.9)	14,930,636	(-6.8)
Treatment costs	16,820,809	15,856,766	(-5.7)	16,485,808	(-2.0)
Total costs	32,843,608	32,991,471	(+0.5)	31,416,445	(-4.3)

CIN = cervical intraepithelial neoplasia; CeCa = cervical cancer; LYs = life years; QALYs = quality-adjusted life years.

and QALYs gained. Still, when all women switched it was effective and cost-saving to offer self-sampling (**Tables 2** and **3**).

Table 3. Discounted simulated effects and costs (both 3% per year) of providing non-attendees with a self-sampling test in a primary HPV screening program under the base-case scenarios, per 100,000 simulated women.

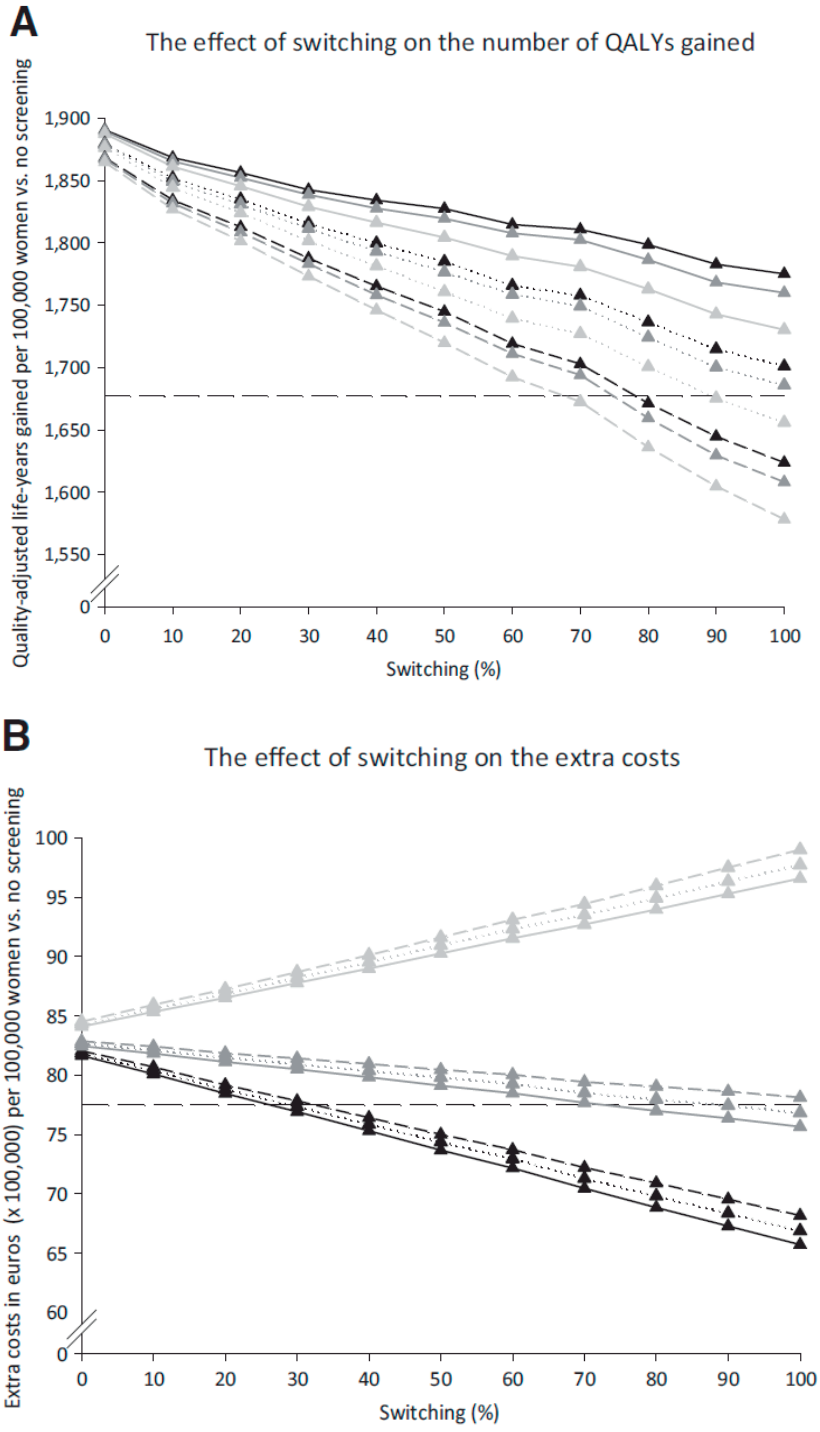
	Base-case scenario, no switching (% vs. no self-sampling)	**Base-case scenario, 100% switching (% vs. no self-sampling)**
LYs gained	1,746 (+12.1)	1,573 (+1.0)
QALYs gained	1,880 (+12.1)	1,701 (+1.4)
Costs in €	8,184,676 (+5.5)	6,687,767 (-13.8)
ICER: Costs in € per LY gained	2,276	Cost-saving
ICER: Costs in € per QALY gained	2,115	Cost-saving

Cost-saving = Cervical cancer screening was both more effective and less costly with versus without offering HPV self-sampling test to non-attendees. LYs = life years; QALYs = quality-adjusted life years;
ICER = incremental cost-effectiveness ratio.

Multivariate sensitivity analyses

Without switching, a decrease in the CIN 2+ sensitivity of self-sampling mainly resulted in fewer QALYs gained (**Figure 3A**), whereas a decrease in the CIN 2+ specificity mainly resulted in increased costs (**Figure 3B**). Both resulted in a higher ICER (**Figure 3C**). However, even when the relative sensitivity and specificity were inferior to that of office-based sampling (i.e. 0.89-0.91 and 0.84-0.85, respectively), QALYs were gained and the ICER was below the threshold of €20,000 per QALY gained, if no women switched.

In all scenarios, switching resulted in fewer QALYs gained (**Figure 3A**). This effect was larger in case the relative sensitivity was lower than 1.00. However, even when the test characteristics of self-sampling were inferior to that of office-based sampling, QALYs were only lost when more than 60% of the women switched (**Table 4**). When they were slightly inferior (i.e. 0.95-0.97 relative sensitivity and 0.94-0.95 relative specificity) or similar (i.e. 1.01-1.02 relative sensitivity and 0.99-1.00 relative specificity), it was effective under every switching scenario. If self-sampling specificity was inferior, the costs of offering HPV self-sampling increased with increasing percentages of switching (**Figure 3B**). Considering a cost-effectiveness threshold of 20,000 per QALY gained, the switching limit was up to 30 percentage points lower (**Table 5**). Therefore, offering self-sampling was not effective or cost-effective when more than 40% of the women switch and test characteristics of self-sampling were inferior to those of office-based sampling.


A
The effect of switching on the number of QALYs gained
Quality-adjusted life-years gained per 100,000 women vs. no screening
1,900
1,850
1,800
1,750
1,700
1,650
1,600
1,550
0
0
10
20
30
40
50
60
70
80
90
100
Switching (%)
B
The effect of switching on the extra costs
Extra costs in euros (x 100,000) per 100,000 women vs. no screening
100
95
90
85
80
75
70
65
60
0
0
10
20
30
40
50
60
70
80
90
100
Switching (%)

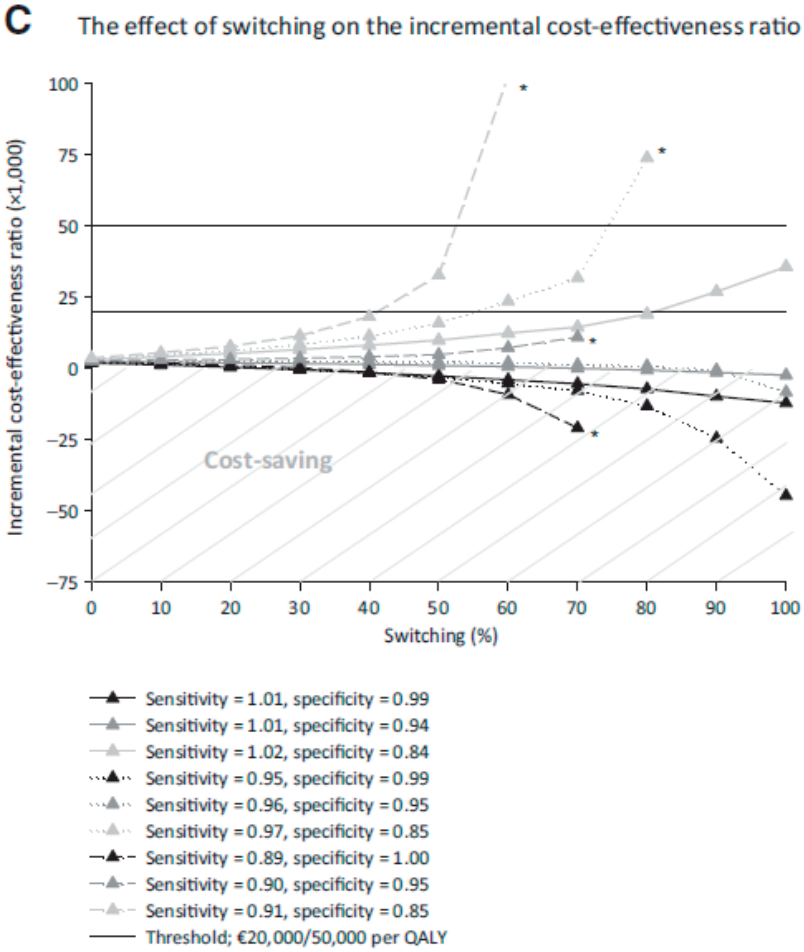


Figure 3. The effect of switching and the relative CIN 2+ sensitivity and specificity on the number of QALYs gained (A), extra costs (B), and ICER (C). Results are given per 100,000 simulated women (3% discounting for costs and effects). The relative CIN 2+ sensitivity and specificity (self-sampling vs. office-based sampling) are indicated by the sensitivity and specificity in the legend.

(C), the combined effect of sensitivity, specificity, and switching on the ICER is only shown when adding a self-sampling test resulted in a gain of QALYs as compared with primary HPV screening alone. Therefore, a negative ICER (i.e., cost-saving) is also dominating (i.e., primary HPV screening with offering a self-sampling test to non-attendees was both more effective and less costly than primary HPV screening alone).

* Beyond this level of switching, offering a self-sampling test resulted in a loss of QALYs as compared with primary HPV screening alone.

Black dashed line = primary HPV screening without offering self-sampling to non-attendees.

Table 4. For what switching percentage does offering self-sampling to non-attendees lead to a loss in QALYs?

Relative CIN2+ sensitivity and specificity (self versus office-based sampling)		**Base case**	**Utility loss due to cytology triage**		**Extra attendance (percentage points)**		**Background risk of higher primary risk women**[a]	**% Extra attendees consisting of higher primary risk women**	
Sensitivity	Specificity		0·000[b]	0·0012	10	3	Average[c]	50	0
1.01	0.99	*Ind.*	*Ind.*	*Ind.*	*Ind.*	*Ind.*	*Ind.*	*Ind.*	>60
1.01	0.94	*Ind.*	*Ind.*	*Ind.*	*Ind.*	>90	*Ind.*	*Ind.*	>50
1.02	0.84	*Ind.*	*Ind.*	*Ind.*	*Ind.*	>70	>80	*Ind.*	>40
0.95	0.99	*Ind.*	*Ind.*	*Ind.*	*Ind.*	>50	>60	*Ind.*	>30
0.96	0.95	*Ind.*	*Ind.*	>90	*Ind.*	>50	>50	*Ind.*	>30
0.97	0.85	>80	*Ind.*	>70	*Ind.*	>40	>40	*Ind.*	>20
0.89	1.00	>70	>70	>70	*Ind.*	>40	>40	*Ind.*	>20
0.90	0.95	>70	>70	>70	*Ind.*	>30	>30	>90	>20
0.91	0.85	>60	>80	>50	>90	>30	>30	>80	>10

For every scenario the minimum switching percentage is given under which it is no longer effective (i.e. QALYs are lost) to offer self-sampling to non-attendees. The switching percentage varies between 0 (i.e. even when no women switch it is not effective to offer self-sampling), >90 (i.e. when more than 90% of the women switch it is not effective to offer self-sampling) to independent (ind.) (i.e. independent of how many women switch, it is always effective to offer self-sampling).

Base-case assumptions: Utility loss due to cytology triage = 0.006 per week, extra attendance = 6 percentage points, background risk of higher primary risk women as compared to the rest of the screen population = 1.7 times higher, % of higher primary risk women in extra attendees = 29%. These variables, if not varied, were held constant at their base-case level.

[a] Background risk for developing cervical cancer.

[b] Since a higher utility loss for primary than triage testing does not seem realistic we also assume no utility loss for primary testing.

[c] Average = Equal to the rest of the screen population.

The effect of the level of utility loss associated with cytology triage was negligible (**Table 4**). Varying the extra attendance or background risk of higher primary risk women had more influence. When the extra attendance was halved (from 6 to 3 percentage points) or if higher primary risk women did not have an elevated background risk, QALYs were lost when more than 50% of the women switched and test characteristics were slightly inferior. When they were inferior, it was no longer effective if more than 30% of the women switched. The most influential parameter was the attendance of higher primary risk women. When they did not attend, it was not effective to offer self-sampling when more than 60% of the women switched and test characteristics were equal. In case they were inferior, this threshold decreased to 10%. For offering self-sampling to be cost-effective, these switching thresholds were even lower (**Table 5**).

Table 5. For what switching percentage is offering self-sampling to non-attendees not (cost-)effective?

Relative CIN2+ sensitivity and specificity (self versus office-based sampling)		Base case	Utility loss due to cytology triage		Costs of self-sampling kit in €		Extra attendance (percentage points)		Background risk of higher primary risk women[a]	% Extra attendees consisting of higher primary risk women	
Sensitivity	Specificity		0.000[b]	0.012	3.50	10	10	3	Average[c]	50	0
1.01	0.99	*Ind.*	*Ind.*	*Ind.*	*Ind.*	*Ind.*	*Ind.*	*Ind.*	*Ind.*	*Ind.*	>60
1.01	0.94	*Ind.*	*Ind.*	*Ind.*	*Ind.*	*Ind.*	*Ind.*	>90	*Ind.*	*Ind.*	>20
1.02	0.84	*>80*	*Ind.*	*>60*	*>90*	*>60*	*Ind.*	>40	>30	*Ind.*	>0
0.95	0.99	*Ind.*	*Ind.*	*Ind.*	*Ind.*	*Ind.*	*Ind.*	>50	>60	*Ind.*	>20
0.96	0.95	*Ind.*	*Ind.*	>90	*Ind.*	*>80*	*Ind.*	>50	>40	*Ind.*	>10
0.97	0.85	>50	*>70*	>40	>60	>40	*>80*	>20	>20	*>80*	>0
0.89	1.00	>70	>70	>70	>70	>70	*Ind.*	>40	>30	*Ind.*	>10
0.90	0.95	>70	>70	>60	>70	>50	*Ind.*	>30	>20	>90	>0
0.91	0.85	>40	>50	>30	>40	>30	>60	>20	>10	>50	>0

For every scenario the minimum switching percentage is given under which it is no longer effective (i.e. QALYs are lost) and/or cost-effective (i.e. €20,000 per QALY gained) to offer self-sampling to non-attendees. The switching percentage varies between 0 (i.e. even when no women switch it is not effective nor cost-effective to offer self-sampling), >90 (i.e. when more than 90% of the women switch it is not effective nor cost-effective to offer self-sampling) to independent (ind.) (i.e. independent of how many women switch, it is always effective and cost-effective to offer self-sampling).

Base-case assumptions: Utility loss due to cytology triage = 0.006 per week, extra attendance = 6 percentage points, background risk of higher primary risk women as compared to the rest of the screen population = 1.7 times higher, % of higher primary risk women in extra attendees = 29%. These variables, if not varied, were held constant at their base-case level.

[a] Background risk for developing cervical cancer.

[b] Since a higher utility loss for primary than triage testing does not seem realistic we also assume no utility loss for primary testing.

[c] Average = Equal to the rest of the screening population.

DISCUSSION

The number of QALYs gained by offering HPV self-sampling to non-attendees was influenced by self-sampling test characteristics, the extra attendance via self-sampling, and the risk of extra attendees. When none of the regular attendees switched to self-sampling, it was always effective to offer HPV self-sampling. Switching resulted in fewer QALYs gained because the probability of being lost to follow-up after a positive self-sampling test was higher than after a positive office-based sampling test. If in addition the sensitivity of self-sampling was lower than that of office-based sampling, the number of QALYs gained decreased even more. However, even when test characteristics were inferior, up to 60% of the regular attendees could switch before the QALYs gained by the 6 percentage points extra attendance were annulled by the QALYs lost by switching. This percentage dropped to 30% when the extra

attendance halved from 6 to 3 percentage points or when higher primary risk women did not have an elevated background risk. It dropped to 10% if higher primary risk women did not attend self-sampling. When also considering a cost-effectiveness threshold of €20,000 per QALY gained, these switching thresholds were 10 to 20 percentage points lower.

Our base-case assumption of 6 percentage points extra attendance was based on the Dutch PROHTECT trial in which a self-sampling kit was sent to 85% of all non-attendees (i.e. the remaining 15% opted-out via a letter).[9] Using another strategy will probably result in another extra attendance rate. If this rate will be lower than 3 percentage points (almost) no women can switch before more QALYs are lost than gained.

We assumed that a subset of the unscreened women have a 1.7 times higher background risk of cervical cancer (i.e. higher primary risk women) than the rest of the screening population, which was based on model calibration. Dugué and colleagues' results have shown that non-attendees of cervical cancer screening (i.e. no cervical smear taken in the past 8 years) had a 3.8-fold increased risk of dying from non-cervical (i.e. non-screened) HPV-associated cancers,[20] which seems to confirm our assumption that at least part of the non-attendees have an increased background risk. Although the PROHTECT study showed that unscreened women (i.e. invited for screening at least once but never attended) attended via self-sampling,[21] it is uncertain whether this is the subset with an increased background risk. If these higher primary risk women do not attend via self-sampling, 10% to 60% of the women can switch before QALYs are lost by offering HPV self-sampling to non-attendees.

The relative sensitivity and specificity of self-sampling as compared with office-based sampling will depend on the type of HPV DNA test used.[15,22] However, even when a validated PCR is used (e.g. GP5+/6+ or the real-time hrHPV Test), it is possible that the sensitivity and specificity of self-sampling are both inferior to that of office-based sampling. In fact, relative test characteristics of self-sampling might even be worse than we assumed in our sensitivity analyses.[15] In that case, the maximum percentage of women that can switch before QALYs are lost is also lower.

Studies in Sweden[23], Finland[24], the United Kingdom[25], and Italy[26] have also shown that offering self-sampling to non-attendees increased screening participation rates. We expect that our conclusions to a large extent apply to other countries and regions with well-organized invitational screening programs with a high compliance and an optimal age range and screening frequency. Even if this would mean that HPV self-sampling would be offered to non-attendees of a primary cytology instead of a primary HPV program. For countries and regions with a lower background risk and/or a more intensive screening program as compared to the Netherlands, benefits of increased participation due to self-sampling are probably lower. In countries without a highly organized invitational program, it may not be feasible to offer a self-sampling test to unscreened women. Instead, it could be offered to the general population by selling it over the counter. However, when screening is not reimbursed by the government, it is questionable to what extent unscreened women will use self-sampling. Indeed, a discrete choice experiment in the US showed that vulnerable adults valued costs higher than the kind of screening offered or the travel distance to obtain

screening.[27] When non-attendance is driven by other factors than feeling uncomfortable or having little time (i.e. factors that can be overcome by using self-sampling instead of going to the clinician,[28] the success of offering self-sampling may be limited.

To our knowledge, this is the first study on the harms and benefits of providing a self-sampling test to non-attendees of a cervical cancer screening program. One of our key assumptions (i.e. extra attendance via self-sampling) was based on observations from the PROHTECT trials.[9,10,21] We extensively studied the effect of the level of switching in combination with the test characteristics of self-sampling and the background cervical cancer risk of its users, which were important and uncertain parameters for the effectiveness of offering self-sampling.

A limitation of our study is that we only focused on unvaccinated women. Screening programs will probably be adapted when vaccinated cohorts reach the start age of screening. A separate analysis for this future situation is beyond the scope of the present analysis. However, we expect that offering self-sampling to non-attendees will be less (cost-)effective, because we expect that fewer health effects can be gained by increasing attendance because of a lower background risk. Another drawback is the limited transposability to other health systems. We expect lower benefits of increased participation due to self-sampling in screening programs that are more intensive than the Dutch future program will be (i.e. 5 lifetime screens at ages 30, 35, 40, 50, and 60 years). Moreover, we might have overestimated the colposcopy compliance after a positive self-sampling test, as this may be lower than after a positive office-based sampling test. This may have resulted in a slight overestimation of the effectiveness of self-sampling. In addition, the relative CIN 2+ specificity as described in our study will be somewhat higher when regular attendees switch, as the prevalence of high-risk HPV infections in women without CIN 2+ is slightly lower in regular attendees as compared with non-attendees attending self-sampling. Furthermore, we did not account for other healthcare that women may get while attending clinic-based screening. This may have underestimated health losses in regular attendees switching to self-sampling, as well as health gains in the small group of extra attendees with a positive self-sampling test complying with their triage invitation.

Offering self-sampling to non-attendees clearly offers an opportunity to increase health benefits in cervical cancer screening if health providers make sure that (1) the relative CIN 2+ sensitivity is at least 0.95, (2) unscreened attendees are recruited with self-sampling, and (3) the total attendance increases by at least 6 percentage points. Otherwise, switching of regular attendees to self-sampling may annul the benefits of self-sampling and even decrease the effectiveness of a primary HPV screening program.

ACKNOWLEDGMENTS

This study was funded by the Dutch National Institute for Public Health and the Environment (007/12 V&Z NvdV/EM).

REFERENCES

1. Netherlands Cancer Registry. Incidence and mortality of cervical cancer; 2012. [cited 2014 Dec 11]. Available from: http://cijfersoverkanker.nl/selecties/dataset_1/img51e671de6dab7.
2. van der Aa MA, Pukkala E, Coebergh JW, et al. Mass screening programmes and trends in cervical cancer in Finland and the Netherlands. Int J Cancer 2008;122:1854-8.
3. Schippers EI. Improvement of the cervical cancer screening program (In Dutch) [Letter of the ministry of Welfare, Public Health and Sport, 17 October 2013 (156231-111126-PG)].
4. Arbyn M, Ronco G, Anttila A, et al. Evidence regarding human papillomavirus testing in secondary prevention of cervical cancer. Vaccine 2012;30 Suppl 5:F88-99.
5. Nobbenhuis MA, Helmerhorst TJ, van den Brule AJ, et al. Primary screening for high risk HPV by home obtained cervicovaginal lavage is an alternative screening tool for unscreened women. J Clin Pathol 2002;55:435-9.
6. Brink AA, Meijer CJ, Wiegerinck MA, et al. High concordance of results of testing for human papillomavirus in cervicovaginal samples collected by two methods, with comparison of a novel self-sampling device to a conventional endocervical brush. J Clin Microbiol 2006;44:2518-23.
7. Casparie M, Tiebosch AT, Burger G, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol 2007;29:19-24.
8. Bos AB, Rebolj M, Habbema JD, et al. Nonattendance is still the main limitation for the effectiveness of screening for cervical cancer in the Netherlands. Int J Cancer 2006;119:2372-5.
9. Gök M, Heideman DA, van Kemenade FJ, et al. HPV testing on self collected cervicovaginal lavage specimens as screening method for women who do not attend cervical screening: cohort study. BMJ 2010;340:c1040.
10. Gök M, van Kemenade FJ, Heideman DA, et al. Experience with high-risk human papillomavirus testing on vaginal brush-based self-samples of non-attendees of the cervical screening program. Int J Cancer 2012;130:1128-35.
11. Habbema JD, van Oortmarssen GJ, Lubbe JT, et al. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed 1985;20:79-93.
12. Health Council of the Netherlands. Population screening for cervical cancer. The Hague: Health Council of the Netherlands, 2011; publication no. 2011/07E.ISBN 978-90-5549-866-6.
13. Verhoef VM, Bosgraaf RP, van Kemenade FJ, et al. Triage by methylation-marker testing versus cytology in women who test HPV-positive on self-collected cervicovaginal specimens (PROHTECT-3): a randomised controlled non-inferiority trial. Lancet Oncol 2014;15:315-22.
14. Meijer CJ, Berkhof J, Castle PE, et al. Guidelines for human papillomavirus DNA test requirements for primary cervical cancer screening in women 30 years and older. Int J Cancer 2009;124:516-20.
15. Arbyn M, Verdoodt F, Snijders PJ, et al. Accuracy of human papillomavirus testing on self-collected versus clinician-collected samples: a meta-analysis. Lancet Oncol 2014.
16. Goldie SJ, Kohli M, Grima D, et al. Projected clinical benefits and cost-effectiveness of a human papillomavirus 16/18 vaccine. J Natl Cancer Inst 2004;96:604-15.
17. Van Ballegooijen M, Rebolj M, Essink-Bot ML, Meerding WJ, Berkers LM, Habbema JDF. De effecten en kosten van het bevolkingsonderzoek naar baarmoederhalskanker in Nederland na de herstructurering. Erasmus MC, afdeling Maatschappelijke Gezondheidszorg, 2006.
18. van den Akker-van Marle ME, van Ballegooijen M, van Oortmarssen GJ, et al. Cost-effectiveness of cervical cancer screening: comparison of screening policies. J Natl Cancer Inst 2002;94:193-204.

19. Eichler HG, Kong SX, Gerth WC, et al. Use of cost-effectiveness analysis in health-care resource allocation decision-making: how are cost-effectiveness thresholds expected to emerge? Value Health 2004;7:518-28.
20. Dugué PA, Lynge E, Rebolj M. Mortality of non-participants in cervical screening: Register-based cohort study. Int J Cancer 2013.
21. Gök M, Heideman DA, van Kemenade FJ, et al. Offering self-sampling for human papillomavirus testing to non-attendees of the cervical screening programme: Characteristics of the responders. Eur J Cancer 2012;48:1799-808.
22. Snijders PJ, Verhoef VM, Arbyn M, et al. High-risk HPV testing on self-sampled versus clinician-collected specimens: a review on the clinical accuracy and impact on population attendance in cervical cancer screening. Int J Cancer 2013;132:2223-36.
23. Broberg G, Gyrd-Hansen D, Miao Jonasson J, et al. Increasing participation in cervical cancer screening: offering a HPV self-test to long-term non-attendees as part of RACOMIP, a Swedish randomized controlled trial. Int J Cancer 2014;134:2223-30.
24. Virtanen A, Nieminen P, Luostarinen T, et al. Self-sample HPV tests as an intervention for nonattendees of cervical cancer screening in Finland: a randomized trial. Cancer Epidemiol Biomarkers Prev 2011;20:1960-9.
25. Szarewski A, Cadman L, Mesher D, et al. HPV self-sampling as an alternative strategy in non-attenders for cervical screening - a randomised controlled trial. Br J Cancer 2011;104:915-20.
26. Giorgi Rossi P, Marsili LM, Camilloni L, et al. The effect of self-sampled HPV testing on participation to cervical cancer screening in Italy: a randomised controlled trial (ISRCTN96071600). Br J Cancer 2011;104:248-54.
27. Pignone M. Understanding preferences for CRC screening programs among vulnerable adults in rural North-Carolina: a discrete choice experiment. Paper presented at 35th annual North American meeting of the Society for Medical Decision Making; October 2013; Baltimore, MD.
28. Bosgraaf RP, Ketelaars PJ, Verhoef VM, et al. Reasons for non-attendance to cervical screening and preferences for HPV self-sampling in Dutch women. Prev Med 2014;64:108-13.

Chapter 7

Cost-effectiveness of a multitarget stool DNA test for colorectal cancer screening of Medicare beneficiaries

Steffie K. Naber*, Amy B. Knudsen*, Ann G. Zauber, Carolyn M. Rutter, Sara E. Fischer, Chester J. Pabiniak, Brittany Soto, Karen M. Kuntz, Iris Lansdorp-Vogelaar

Submitted.

*Shared first authorship.

ABSTRACT

Background: The Centers for Medicare and Medicaid Services (CMS) recently began covering a multitarget stool DNA (mtSDNA) test for colorectal cancer (CRC) screening of Medicare beneficiaries.
Objective: To evaluate whether mtSDNA testing is a cost-effective alternative to other CRC screening strategies reimbursed by CMS, and if not, under what conditions it could be.
Design: 3 microsimulation models.
Data sources: Published literature.
Target population: Previously unscreened 65-year-olds.
Time horizon: Lifetime.
Interventions: No screening, triennial mtSDNA testing, and 6 other screening strategies reimbursed by CMS.
Outcome measures: Discounted life years gained (LYG) and lifetime costs (CMS perspective), threshold reimbursement rates, and threshold adherence rates. Outcomes are expressed as the median of the 3 models.
Results of base-case analysis: Compared to no screening, triennial mtSDNA screening resulted in 82 LYG per 1,000 simulated individuals. This was more than for 5-yearly sigmoidoscopy (80 LYG), but fewer than for every other screening strategy reimbursed by CMS (88 to 103 LYG). At its current reimbursement rate of $493, mtSDNA was the most expensive strategy. Per-test reimbursement would need to be below $33 for triennial screening to be an efficient and potentially cost-effective screening option. Even if adherence were 30% higher than with other strategies (90% versus 69%), triennial mtSDNA screening would not be cost-effective at its current reimbursement rate.
Results of sensitivity analysis: Per-test reimbursement rates of $47 and $58 could be supported with biennial or annual mtSDNA screening, respectively.
Limitations: The models assume that, conditional on true disease status, test performance does not vary across repeat screens.
Conclusions: Triennial mtSDNA screening is less effective than nearly all other CRC screening tests reimbursed by CMS. At its current reimbursement rate, it also has higher costs than all other strategies, making it an inefficient screening option.

INTRODUCTION

Colorectal cancer (CRC) is the second most common cause of cancer death in the United States.[1] Randomized trials of fecal occult blood tests (FOBTs) and flexible sigmoidoscopy have shown that screening can effectively reduce both CRC incidence[2-6] and mortality[2-12]. In June 2016, the US Preventive Services Task Force updated their CRC screening recommendations, including guidelines on the use of FOBTs, flexible sigmoidoscopy, colonoscopy, and the recently developed multitarget stool DNA test (mtSDNA).[13]

In April 2014, Imperiale et al.[14] published the findings of a study evaluating the test performance of the mtSDNA test, Cologuard® (Exact Sciences Corporation), based on a single round of screening. Cologuard combines DNA assays for multiple aberrant gene mutations and a proprietary fecal immunochemical assay. Compared with a fecal immunochemical test (FIT), Cologuard demonstrated higher sensitivity for CRC and advanced adenomas but lower specificity. In October 2014, the Centers for Medicare and Medicaid Services (CMS) granted coverage of Cologuard once every 3 years (the interval recommended by the manufacturer) for asymptomatic, average-risk Medicare beneficiaries.[15] Following rules for reimbursement of laboratory tests, the reimbursement rates for the various components of Cologuard were summed to determine its total reimbursement rate of $492.72 per test. CMS requested an analysis of mtSDNA screening of Medicare enrollees from the MITRE Corporation. MITRE commissioned investigators from the Cancer Intervention and Surveillance Modeling Network (CISNET) to assess whether mtSDNA testing is a cost-effective alternative to other CRC screening strategies available to Medicare beneficiaries, and if not, to assess at what reimbursement rate, level of screening uptake, or screening interval it could be a cost-effective option.

METHODS

CISNET models

We used 3 independently-developed microsimulation models of CRC from the National Cancer Institute's CISNET consortium—the CRC Simulated Population Model for Incidence and Natural History (CRC-SPIN), Microsimulation Screening Analysis for CRC (MISCAN), and Simulation Model of CRC (SimCRC)—to evaluate the cost-effectiveness of screening Medicare beneficiaries for CRC with the mtSDNA test. All models describe the natural history of CRC in an unscreened population, based on the adenoma-carcinoma sequence.[16-18] Simulated persons enter free of colonic and rectal lesions at age 20. As they age, they are at risk of developing adenomas. Each adenoma may grow in size, and some may transition to a preclinical (i.e., undiagnosed) CRC. Preclinical cancers may progress in stage, and some may become symptomatic, at which point the person becomes a clinically-detected case. Persons may die from causes other than CRC at any age, and persons with clinically-detected CRC may die from the disease.

Each model has a screening component that allows the natural history of CRC to be interrupted due to the detection of a preclinical cancer or the detection and removal of adenoma(s). With screening, a simulated person with an underlying lesion has a chance of having it detected depending on the sensitivity of the screening test and, for endoscopic tests, whether the lesion is within the reach of the scope. Screened persons without an underlying lesion may have a false-positive test result and undergo an unnecessary follow-up colonoscopy. Non-adenomatous polyps (e.g., hyperplastic polyps) are not modeled explicitly, but their detection is reflected in the false-positive rates of the tests. The impact of screening depends on the characteristics of the test performed, and on how frequently it is repeated.

CRC screening

We used the models to estimate life expectancy and lifetime costs among a cohort of previously unscreened 65-year-olds in the absence of CRC screening, with mtSDNA every 3 years (as specified in the final coverage determination) and with 6 other strategies included in CRC screening recommendations[13] and available to Medicare beneficiaries[19]: annual fecal occult blood testing (FOBT) with either a high-sensitivity guaiac-based FOBT (gFOBT) or a FIT, 5-yearly flexible sigmoidoscopy, 10-yearly flexible sigmoidoscopy with annual gFOBT or annual FIT, and 10-yearly colonoscopy. We assumed all screening begins at age 65 and ends no later than age 75. Individuals with a positive non-colonoscopy screening test undergo a follow-up colonoscopy. Individuals with adenomas detected at a screening or follow-up colonoscopy transitioned to an adenoma surveillance regimen,[20] with colonoscopy performed every 3 or 5 years (dependent on findings) until at least age 85.

For the base-case analysis, we assumed 100% adherence to all screening, follow-up, and surveillance procedures. Alternative assumptions were explored in a sensitivity analysis.

For all tests, sensitivity, specificity and reach estimates were based on literature (**Table 1**). The risks of complications from colonoscopy were obtained from a study by van Hees et al.[21] that estimated excess risks of serious gastrointestinal events, other gastrointestinal events, and cardiovascular events by age and polypectomy status among Medicare beneficiaries undergoing colonoscopy compared with a matched control group that did not have colonoscopy.[22]

Costs

The analysis was conducted from the CMS perspective, and as such, costs were valued by Medicare reimbursement rates and excluded beneficiary copayments and cost-sharing payments. Screening costs were generally based on 2014 average Medicare payments (**Table 2**). For a detailed description of the derivation of these costs, see the **Appendix**.

Net costs of CRC-related care by stage at diagnosis and phase of care were obtained from an analysis of 1998-2003 SEER-Medicare linked data[23] (personal communication, Robin Yabroff, PhD and Martin Brown, PhD) and were updated to 2014 dollars using the Consumer Price Index (**Table 2**). Costs from that period do not reflect the use of the expensive monoclonal antibodies cetuximab and bevacizumab, which received FDA approval for treatment of CRC in 2004.[24,25] Higher costs of care were therefore explored in a sensitivity analysis.

Table 1. Test characteristics[a] used in the analysis.

Screening Test Test Characteristic	Base-Case Value, %	Source	Sensitivity Analysis		
			Worst-Case Value, %	Best-Case Value, %	Source
mtSDNA		Imperiale et al, 2014[14]			Imperiale et al, 2014[14]
Specificity[b]	89.8		Not varied	Not varied	Not varied
Sensitivity for adenomas ≤9 mm	17.2[c]		15.9[c]	18.6[c]	
Sensitivity for adenomas ≥10 mm	42.4[d]		38.7[d]	46.2[d]	
Sensitivity for colorectal cancer	92.3		84	97	
FIT (cutoff 20 µg of hemoglobin per g of feces)		Imperiale et al, 2014[14]			Imperiale et al, 2014[14]
Specificity[b]	96.4		Not varied	Not varied	Not varied
Sensitivity for adenomas ≤9 mm	7.6[c]		6.7[c]	8.6[c]	
Sensitivity for adenomas ≥10 mm	23.8[d]		20.8[d]	27[d]	
Sensitivity for colorectal cancer	73.8		62.3	83.3	
gFOBT		Zauber et al, 2008[46]			
Specificity[b]	92.5		Not varied	Not varied	Not varied
Sensitivity for adenomas 1-5 mm	7.5[e]		7.5[e]	7.5[e]	Zauber et al, 2008[46]
Sensitivity for adenomas 6-9 mm	12.4		10	26.2	Zauber et al, 2008[46]
Sensitivity for adenomas ≥10 mm	23.9		17.7	49.4	Zauber et al, 2008[46]
Sensitivity for colorectal cancer	70		61.5	79.4	Levi et al, 2011[47] Allison et al, 1996[48]
Colonoscopy (within reach)[f]					
Specificity[b]	86[g]	Schroy et al, 2013[49]	Not varied	Not varied	Not varied
Sensitivity for adenomas 1-5 mm	75	van Rijn et al, 2006[50]	70	79	Zauber et al, 2008[51]
Sensitivity for adenomas 6-9 mm	85	van Rijn et al, 2006[50]	80	92	Zauber et al, 2008[51]
Sensitivity for adenomas ≥10 mm	95	van Rijn et al, 2006[50]	93.1	99.5	Johnson et al, 2008[52]

Table 1 (continued). Test characteristics[a] used in the analysis.

Screening Test Test Characteristic	Base-Case Value, %	Source	Sensitivity Analysis		
			Worst-Case Value, %	Best-Case Value, %	Source
Sensitivity for colorectal cancer	95	By assumption	93.1	99.5	By assumption
Reach[h]	95 to end of cecum, remainder between rectum and cecum	By assumption	Not varied	Not varied	
Sigmoidoscopy (within reach)					
Specificity[b]	87[g]	Weissfeld et al, 2005[53]	Not varied	Not varied	
Sensitivity for adenomas 1-5 mm	75	By assumption	70	79	
Sensitivity for adenomas 6-9 mm	85	By assumption	80	92	
Sensitivity for adenomas ≥10 mm	95	By assumption	93.1	99.5	
Sensitivity for colorectal cancer	95	By assumption	93.1	99.5	
Reach	76-88 to sigmoid-descending junction; 0 beyond the splenic flexure	Atkin et al, 2002[54] Painter et al, 1999[55]	Not varied	Not varied	

FIT = fecal immunochemical test; gFOBT = sensitive guaiac-based fecal occult blood test; mtSDNA = multitarget stool DNA test.

[a] Sensitivity estimates are per person for stool-based tests and per-lesion for endoscopic tests. Specificity is defined as the probability of a negative test result among persons who do not have adenomas or colorectal cancer.

[b] Specificity is defined as the probability of a negative test result among persons who do not have adenomas or colorectal cancer.

[c] Sensitivity for persons with non-advanced adenomas. For persons with ≤5 mm adenomas, we assume that the sensitivity of the test is equal to the positivity rate in persons without adenomas (i.e., 1 – specificity). The sensitivity for persons with 6-9 mm adenomas is chosen such that the weighted average sensitivity for persons with ≤5 mm and with 6-9 mm adenoma(s) is equal to that of non-advanced adenomas.

[d] Sensitivity for persons with advanced adenomas (i.e., adenomas ≥10 mm and/or adenomas with advanced histology). Sensitivity was not reported for the subset of ≥10 mm adenomas.

[e] We assume that ≤5 mm adenomas do not bleed, and therefore cannot cause a positive stool test. We also assume that gFOBT can be positive due to bleeding from other causes, the probability of which is equal to positivity rate in persons without adenomas (i.e., 1 – 0.925).

[f] We assume the same test characteristics for screening colonoscopies as for colonoscopies for diagnostic follow-up or for surveillance. We assume no correlation in findings between sigmoidoscopy and subsequent diagnostic colonoscopy.

[g] The lack of specificity with endoscopy reflects the detection of non-adenomatous polyps, which, in the case of sigmoidoscopy, may lead to unnecessary diagnostic colonoscopy, and in the case of colonoscopy screening, leads to unnecessary polypectomy, which is associated with an increased risk complications.

[h] We assume that 5% of persons undergoing colonoscopy require 2 procedures to achieve complete visualization and that the cecum is ultimately visualized in 95% of patients.

Cost-effectiveness analysis

We used the simulation models to calculate lifetime costs of CRC screening and related care and life expectancy for a previously unscreened cohort of 65-year-old Medicare beneficiaries under 8 CRC screening strategies, including no screening. We conducted an incremental cost-effectiveness analysis from the perspective of CMS and discounted both future costs and life years 3% annually to account for time preference for present over future outcomes.[26] Screening strategies were ranked by increasing costs. Strategies that were more costly and less effective than another strategy (i.e., strongly dominated strategies) were eliminated from consideration because they were inefficient screening options. Of the remaining strategies, those that were less effective and less costly than another but provided an additional life year gained (LYG) at a higher incremental cost (i.e., weakly dominated strategies) were also eliminated from consideration. The relative performance of the remaining non-dominated strategies was measured using the incremental cost-effectiveness ratio (ICER), defined as the additional cost of a specific strategy, divided by its additional clinical benefit (in this case, LYG), compared with the non-dominated strategy with costs closest to, but lower than, the strategy of interest.

All non-dominated strategies represent the set of potentially cost-effective options and together comprise the efficient frontier. Which strategy is ultimately deemed to be cost-effective depends on the willingness to pay for a LYG. Although there is no official willingness-to-pay threshold in the US, a strategy with an ICER less than $50,000-100,000 per LYG is generally considered to provide a good value.[27] For this analysis we assumed a willingness-to-pay threshold of $100,000 per LYG.

Threshold analyses

If the triennial mtSDNA test strategy was found to be dominated by other screening options, we calculated the maximum cost per mtSDNA test (i.e., the threshold cost) for that strategy to be on the efficient frontier (i.e., to be potentially cost-effective). Since the availability of the mtSDNA test could entice a previously unscreened individual to undergo screening, we also identified the threshold mtSDNA test cost for scenarios in which the adherence of the mtSDNA strategy was greater than that of all other screening strategies. For that analysis we assumed an overall adherence rate of 69% for each test,[28] with this 69% of the population completely adherent to screening and the remainder completely non-adherent. We varied the adherence for the mtSDNA test strategy from 69% to 100%. Subsequently, the threshold mtSDNA test cost was calculated comparing lifetime costs and LYG with mtSDNA testing at these higher adherence rates to competing strategies at an adherence rate of 69%.

Outcomes

Outcomes are reported as the results predicted by each of the three models, focusing on the median prediction, along with estimates from the other two models, which define the range across models.

Table 2. Reimbursement for screening tests and for colonoscopy complications, and annual reimbursements for cancer care used in the base-case and sensitivity analysis.

	REIMBURSEMENT[a] ($)				
Screening tests					
mtSDNA	493				
FIT	22				
gFOBT	4				
Colonoscopy, without polypectomy, by indication					
- Screening	699				
- Diagnostic	591				
- Surveillance	681				
Colonoscopy, with polypectomy	813				
Sigmoidoscopy	274				
Colonoscopy complications					
Serious GI complication (perforations, GI bleeding, transfusions)	6,657				
Other GI complication (paralytic ileus, nausea and vomiting, dehydration, abdominal pain)	4,743				
Cardiovascular complication (myocardial infarction or angina, arrhythmias, congestive heart failure, cardiac or respiratory arrest, syncope, hypotension, or shock)	5,199				
	ANNUAL REIMBURSEMENT OF CANCER CARE ($)				
Phase of care[b] Stage at diagnosis	**Base-case analysis**	**Sensitivity analysis**[c]			
		10% higher	**25% higher**	**50% higher**	**75% higher**
Initial phase					
I	29,100	---------- Not varied ----------			
II	40,159	---------- Not varied ----------			
III	48,965	53,861	61,206	73,447	85,688
IV	63,939	70,333	79,924	95,908	111,893
Continuing phase					
I	2,316	---------- Not varied ----------			
II	2,158	---------- Not varied ----------			
III	3,085	---------- Not varied ----------			
IV	9,562	---------- Not varied ----------			
Terminal phase, non-CRC death					
I	12,853	---------- Not varied ----------			
II	11,242	---------- Not varied ----------			
III	14,873	---------- Not varied ----------			
IV	39,933	---------- Not varied ----------			

Terminal phase, CRC death					
I	52,166	57,383	65,208	78,249	91,291
II	52,019	57,221	65,024	78,028	91,033
III	54,812	60,293	68,515	82,217	95,920
IV	73,562	80,918	91,952	110,343	128,733

CRC = colorectal cancer; FIT = fecal immunochemical test; gFOBT = sensitive guaiac-based fecal occult blood test; GI = gastrointestinal; mtSDNA = multitarget stool DNA test.

[a] Costs of stool-based tests are based on the 2014 (for gFOBT and FIT) and 2015 (for mtSDNA) Clinical Laboratory Fee Schedule. Costs of endoscopic procedures are based on 2014 average payments for a screening sigmoidoscopy and for each type of colonoscopy and include payments for pathology, anesthesia services, and anesthetic agents (i.e., propofol).

[b] The initial phase of care is the first 12 months after diagnosis, the last year of life phase is the final 12 months of life, and the continuing phase is all the months between the initial and last year of life phases.

[c] Reimbursements in the initial phase of care for cases diagnosed at stage III or IV and in terminal phase of care for those who die from colorectal cancer are allowed to vary in the sensitivity analysis. All other reimbursement rates remain at base-case values.

Sensitivity analyses

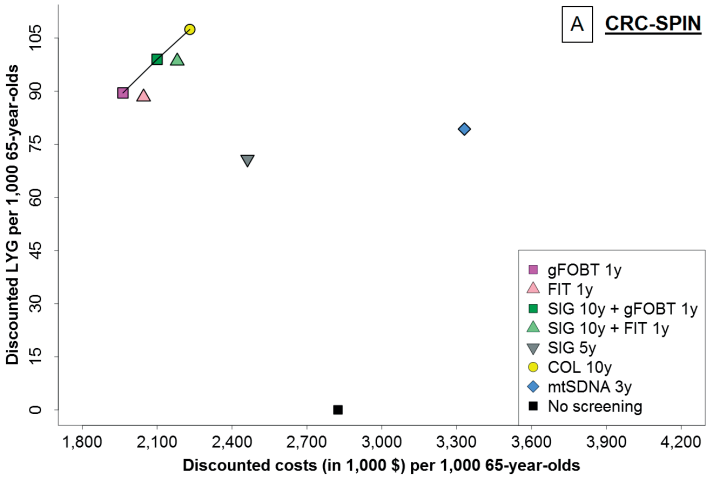
Threshold costs for mtSDNA were also identified in sensitivity analyses with: higher estimates of the reimbursement for cancer care (**Table 2**); higher and lower estimates of the sensitivity of either mtSDNA testing or of all other screening modalities (**Table 1**); and annual and biennial screening intervals for mtSDNA testing.

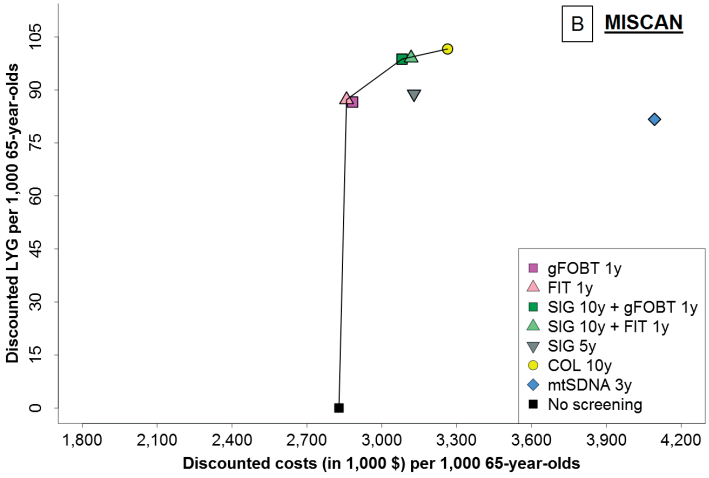
RESULTS

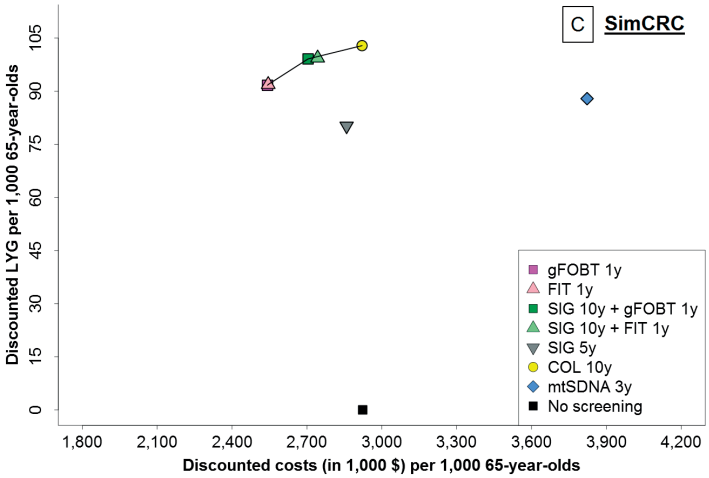
In the absence of screening, 64 (range across models: 61-64) per 1,000 65-year-olds will be diagnosed with CRC in their lifetimes (**Table 3**), resulting in approximately $2.8 million (range: $2.8-2.9 million) in discounted lifetime direct medical costs. All screening strategies yielded large reductions in CRC incidence and mortality. Assuming 100% adherence, the reduction in lifetime risk of CRC with one of the established screening strategies ranged from 50% (range: 36-59%) with annual FIT screening to 73% (range: 58-86%) with 10-yearly colonoscopy screening (**Appendix Figure 1, Panel A**). CRC risk reduction with triennial mtSDNA testing was 46% (range: 33-54%), which was slightly less than that of annual FIT. Reductions in the lifetime risk of CRC death (**Appendix Figure 1, Panel B**) were higher than reductions in incidence but followed a similar pattern. The reduction in lifetime risk of CRC death with triennial mtSDNA testing was nearly identical to that of 5-yearly sigmoidoscopy at 66% (range: 62-68%).

Cost-effectiveness analysis

Three screening strategies were found to be efficient by all models: 10-yearly colonoscopy, 10-yearly sigmoidoscopy with annual gFOBT, and annual FOBT, although the specific FOBT strategy (i.e., gFOBT or FIT) varied across models (**Figure 1, Table 3**). In 1 model (MISCAN),


A
CRC-SPIN
Discounted LYG per 1,000 65-year-olds
0
15
30
45
60
75
90
105
1,800
2,100
2,400
2,700
3,000
3,300
3,600
3,900
4,200
Discounted costs (in 1,000 $) per 1,000 65-year-olds
gFOBT 1y
FIT 1y
SIG 10y + gFOBT 1y
SIG 10y + FIT 1y
SIG 5y
COL 10y
mtSDNA 3y
No screening


B
MISCAN
Discounted LYG per 1,000 65-year-olds
0
15
30
45
60
75
90
105
1,800
2,100
2,400
2,700
3,000
3,300
3,600
3,900
4,200
Discounted costs (in 1,000 $) per 1,000 65-year-olds
gFOBT 1y
FIT 1y
SIG 10y + gFOBT 1y
SIG 10y + FIT 1y
SIG 5y
COL 10y
mtSDNA 3y
No screening


C
SimCRC
Discounted LYG per 1,000 65-year-olds
0
15
30
45
60
75
90
105
1,800
2,100
2,400
2,700
3,000
3,300
3,600
3,900
4,200
Discounted costs (in 1,000 $) per 1,000 65-year-olds
gFOBT 1y
FIT 1y
SIG 10y + gFOBT 1y
SIG 10y + FIT 1y
SIG 5y
COL 10y
mtSDNA 3y
No screening

← **Figure 1.** Discounted costs and discounted life years gained per 1,000 persons aged 65 years for eight colorectal cancer screening strategies and the efficient frontier connecting the economically efficient strategies, for CRC-SPIN (Panel A), MISCAN (Panel B) and SimCRC (Panel C) models.

Discounted costs and life years gained reflect total costs and life years gained of a screening program, accounting for time preference for present over future outcomes. Life years gained are plotted on the *y*-axis, and total costs are plotted on the *x*-axis. Each possible screening strategy is represented by a point. Strategies that form the solid line connecting the points lying left and upward are the economically rational subset of choices. This line is called the *efficient frontier*. The inverse slope of the line represents the incremental cost-effectiveness ratio of the connected strategies. Points lying to the right and beneath the line represent the dominated strategies. Screening with the multitarget stool DNA test every 3 years has higher costs and fewer life years gained than screening annually with either gFOBT or FIT, and the multitarget stool DNA strategy is therefore strongly dominated.

COL = colonoscopy; FIT = fecal immunochemical test; gFOBT = guaiac-based fecal occult blood test; LYG = life years gained; mtSDNA = multitarget stool DNA test; SIG = flexible sigmoidoscopy.

Table 3. Undiscounted colorectal cancer cases and deaths, and discounted costs and life years gained with associated incremental cost-effectiveness ratio of no colorectal cancer screening and seven colorectal cancer screening strategies in a cohort of 1,000 previously unscreened 65-year-olds, by model.

	CRC-SPIN					MISCAN					SimCRC				
Strategy	**CRC cases**	**CRC deaths**	**Lifetime costs,[a] million $**	**LYG[a]**	**ICER, $**	**CRC cases**	**CRC deaths**	**Lifetime costs,[a] million $**	**LYG[a]**	**ICER, $**	**CRC cases**	**CRC deaths**	**Lifetime costs,[a] million $**	**LYG[a]**	**ICER, $**
No screening	64	23	2.824	0	D	61	25	2.828	0	--	64	25	2.924	0	D
gFOBT 1y	25	06	1.963	089.5	--	39	08	2.883	86.6	D	31	07	2.542	91.6	--
FIT 1y	27	06	2.046	088.3	D	39	08	2.859	87.2	300	32	07	2.545	91.9	12,700
SIG 5y	29	09	2.462	070.8	D	30	07	3.129	88.9	D	28	09	2.859	80.1	D
SIG 10y + gFOBT 1y	17	04	2.100	099.0	14,400	29	06	3.081	98.7	19,400	23	05	2.705	99.1	22,100
SIG 10y + FIT 1y	17	04	2.180	098.5	D	29	06	3.118	99.0	D[b]	23	05	2.743	99.3	D[b]
COL 10y	09	02	2.231	107.4	15,500	25	05	3.264	101.6	63,600	17	04	2.921	102.8	57,600
mtSDNA 3y	30	08	3.331	79.3	D	41	09	4.093	81.7	D	34	08	3.823	87.9	D

-- = default strategy (i.e., the least costly and least effective non-dominated strategy); COL = colonoscopy; CRC = colorectal cancer; D = dominated; FIT = fecal immunochemical test; gFOBT = high sensitivity guaiac-based fecal occult blood test; ICER = incremental cost-effectiveness ratio; LYG = life years gained compared with no screening; mtSDNA = multitarget stool DNA test; SIG = flexible sigmoidoscopy.

[a] Future costs and life years are discounted at a 3% annual rate.

[b] Indicates a dominated strategy is weakly dominated (i.e., one of the other strategies provides more life years gained than this strategy, and it has a lower incremental cost-effectiveness ratio). All other dominated strategies are strongly dominated (i.e., provide fewer life years gained and have higher total costs than another strategy).

triennial mtSDNA testing yielded the fewest LYG of all evaluated strategies, and in the other 2 models it had the second-fewest LYG after 5-yearly sigmoidoscopy. All models found that the current reimbursement rate of $492.72 per test made triennial mtSDNA testing the most expensive strategy of those considered. With higher costs and fewer LYG it was not an efficient strategy.

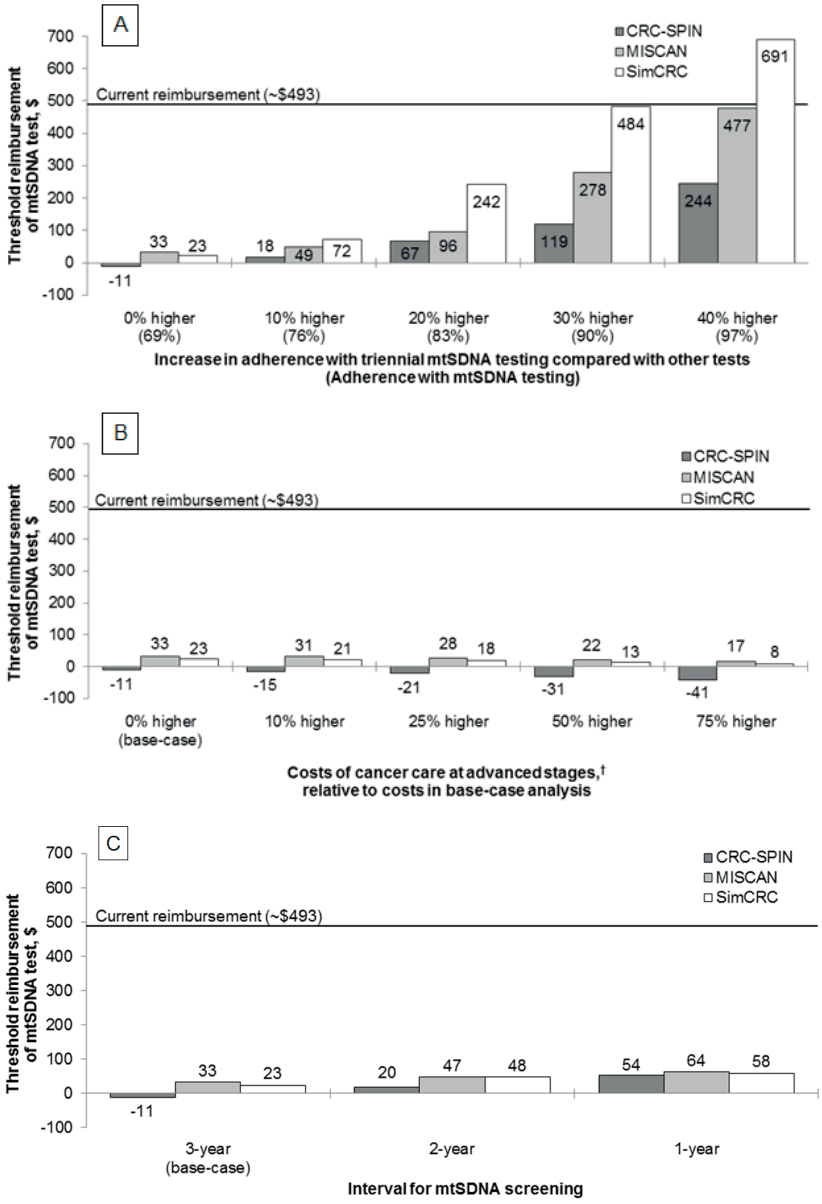


Figure 2. Sensitivity analyses: Reimbursement thresholds for the mtSDNA test at which the mtSDNA test strategy is efficient compared with other reimbursed CRC screening strategies for different levels of adherence with the mtSDNA strategy (Panel A), for different levels of the cost of cancer care (Panel B), and for different intervals of screening with the mtSDNA test (Panel C).

mtSDNA = multitarget stool DNA test.

* The screening interval for the mtSDNA test is every 3 years unless otherwise noted.

† Initial phase of care for cases diagnosed at stage III or IV and terminal phase of care for those who die from CRC, regardless of stage at diagnosis. Cost at all other stages are at base-case levels.

Threshold analyses

In threshold analyses, 2 models (MISCAN and SimCRC) found that the reimbursement for the mtSDNA test must be considerably lower, in the range of $23-33 per test, for triennial mtSDNA screening to be an efficient and potentially cost-effective strategy (**Figure 2**). In one model (CRC-SPIN), there was no level of reimbursement at which triennial mtSDNA testing would be cost-effective compared with currently recommended screening options (i.e., the threshold cost was negative).

If the triennial mtSDNA test strategy would motivate individuals who would not otherwise be screened to participate in screening, then the threshold cost at which the mtSDNA strategy would be on the efficient frontier would increase. Two models (MISCAN and SimCRC) estimate that adherence with mtSDNA testing would need to be 30-41% better than with other tests in order for triennial mtSDNA testing to be efficient at the base-case reimbursement rate of $492.72 (**Figure 2, Panel A**). With these increases, overall adherence with triennial mtSDNA testing would be nearly perfect, at 90-97% (i.e., nearly all eligible persons adherent with all screening, follow-up and surveillance procedures). CRC-SPIN estimates that the current reimbursement rate could not be supported even with 100% adherence to the mtSDNA strategy.

Sensitivity analyses

Increasing the costs of cancer care to reflect the use of targeted treatments resulted in threshold reimbursement rates that were even lower than those from the base-case analysis (**Figure 2, Panel B**). If mtSDNA was assumed to perform better at detecting disease or if other modalities were assumed to perform worse than assumed in the base-case analysis, estimated reimbursement rates increased to $39 (range: $7-47) and $45 (range: $38-51), respectively (**Appendix Figure 2**). When the interval of mtSDNA screening was shortened to 2 years or 1 year, threshold costs increased to $47 (range: $20-48) and $58 (range: $54-64), respectively (**Figure 2, Panel C**).

DISCUSSION

This study showed that despite having superior per-test sensitivity compared to FIT, with perfect adherence, a program of triennial mtSDNA screening was slightly less effective in terms of LYG than a program of annual FIT screening. Triennial mtSDNA screening resulted in 82 LYG per 1,000 65-year-olds (range: 79-88) and, at its current reimbursement rate of $492.72 per test, it had a net cost of $0.9 million (range: $0.5-1.3 million) compared to no screening. The lifetime costs of triennial mtSDNA were higher than all other established screening strategies, and this strategy was therefore not cost-effective compared to other screening options available to Medicare beneficiaries. Per-test reimbursement for the mtSDNA test would need to be less than $33 for triennial screening to be a potentially cost-effective option. Higher reimbursement rates could be supported with more frequent screening: $47

(range: $20-48) per test with biennial screening and $58 (range: $54-64) per test with annual screening.

Despite its higher sensitivity for advanced adenomas and cancer, triennial mtSDNA testing yielded fewer LYG than the two other stool-based CRC screening strategies evaluated. The lower effectiveness can be explained by the longer screening interval. With annual FIT, a person with an advanced adenoma has a 23.8% probability of a positive test result each year due to the advanced adenoma. This amounts to a probability of a positive test due to the advanced adenoma of 55.8% after 3 annual screens [i.e., $1 - (1 - 0.238)^3$], assuming that screening results are independent within an individual, conditional on true clinical state. [Note that this is a simplified example. The models also account for changes in the number and size of lesions as simulated individuals age.] Triennial mtSDNA screening only has 1 opportunity in the 3-year period to yield a positive result due to the advanced adenoma, with the probability equal to 42.4%. For individuals with undiagnosed cancer, the probability of a positive test due to that cancer is 98.2% after 3 annual screens with FIT [i.e., $1 - (1 - 0.738)^3$], versus 92.3% after a single screen with mtSDNA. If performed annually or biennially, programmatic sensitivity of mtSDNA screening would increase, although the probability of a false-positive test result would also increase.

When reimbursed at $492.72 per test, mtSDNA would only be cost-effective compared to other strategies if it would increase adherence to more than 90% of eligible adults. Given current levels of adherence of 69%,[28] such a large increase in adherence is unlikely. The test may appeal to some unscreened persons because it is non-invasive, has higher sensitivity (but lower specificity) than other stool tests and can be performed less frequently. However, it is still a stool test (and in fact, requires patients to sample from the collected stool for the immunochemical assay portion of the test), and, as such, the demonstrated barriers to this form of screening, such as handling of stool and storing stool in the house for a short period of time also apply to mtSDNA testing.[29] Furthermore, mtSDNA does not eliminate the barriers common to all screening tests, namely financial barriers, failure of clinicians to advise about CRC screening, and not knowing testing was necessary.[29]

CMS's high reimbursement rate for Cologuard -- more than 10 times the reimbursement for other stool-based screening tests for CRC -- is the result of federal regulations for setting reimbursement for new diagnostic laboratory tests. Payment for a new diagnostic laboratory test is set by one of two approaches: "cross-walking" or "gap-filling".[30] Cross-walking is used if the new test is comparable to one or more existing tests already reimbursed under the Clinical Laboratory Fee Schedule (CLFS); reimbursement for the new test is set equal to the reimbursement for the comparable test(s). If no comparable test exists, then payment for a new diagnostic laboratory test is set using the gap-filling approach. With that approach, other information is taken into consideration, including information on reimbursement for the test in non-Medicare settings and resource use required for other relevant tests. Reimbursement for Cologuard was set by cross-walking to three existing codes on the CLFS (81315, 81275, and 82274), yielding the 2014 reimbursement of $492.72.[31] Had CMS instead used the gap-filling approach, it is possible that the cost-effectiveness of screening for CRC

with Cologuard compared with other tests reimbursed by CMS could have factored into the reimbursement rate. Reimbursement rates for all clinical laboratory tests, including gFOBT, FIT, and Cologuard, are likely to change when Section 216 of the Protecting Access to Medicare Act of 2014 goes into effect.[32] As of January 2018, reimbursement for such tests will be set based on weighted median private-payer rates.[33]

Our findings are in line with our previous analysis of the cost-effectiveness of stool DNA testing, in which we considered a hypothetical test called "sDNA version 2.0".[34,35] This test had sensitivity and specificity similar to the mtSDNA test, but was never available to the public. For that test, we found threshold reimbursement rates of $17-41, which are similar to the current estimates.

A recent modeling study by Ladabaum and Mannalithara[36] showed that mtSDNA testing is not cost-effective, unless participation rates would be significantly higher than with other screening modalities. Although the study is comparable to ours, results are not directly comparable because screening was simulated from age 50 instead of age 65. Obviously, relative differences in incidence reduction and LYG between screening strategies are larger when considering a broader screening age range. Moreover, the study only evaluated the cost-effectiveness of mtSDNA testing compared to FIT and colonoscopy screening strategies, while we found that annual gFOBT screening (either alone or with 10-yearly sigmoidoscopy) could be cost-effective as well. The authors acknowledge that the reimbursement rate of mtSDNA includes a patient support program, while the reimbursement rate of FIT does not. Although we did not explicitly add costs for patient support to the costs of other screening tests (as they did for FIT), we did estimate the increase in adherence that is required for mtSDNA testing to become a cost-effective alternative. We found that even if the patient support program would lead to an increase in adherence of 30% (90% for mtSDNA versus 69% for other tests), triennial mtSDNA testing would still not be cost-effective compared to other screening options.

An important strength of the current study is the use of 3 independently-developed models. Some limitations are noteworthy. First, the models assume that all CRCs arise through the traditional adenoma–carcinoma sequence, and none incorporate a separate pathway for sessile serrated adenomas. However, for both FIT and mtSDNA, the sensitivity for detecting advanced adenomas includes both traditional advanced and sessile serrated adenomas.[14] Our models would underestimate the effectiveness of mtSDNA compared to FIT only if mtSDNA sensitivity for serrated adenomas is greater than FIT *and* these lesions have higher malignant potential than traditional advanced adenomas. There is evidence to suggest that FIT might be less sensitive than mtSDNA for sessile serrated adenomas,[14,37] but their malignant potential seems comparable to that of traditional advanced adenomas.[38]

Second, the models simulate the progression from adenoma to CRC by allowing adenomas to increase in size over time. Because adenoma size and the presence of villous components or high-grade dysplasia are highly correlated,[39] size indirectly represents histology and grade. However, none of the models separately simulate the step from adenoma with low-grade dysplasia to an adenoma with high-grade dysplasia. For the sensitivity of

mtSDNA to detect large adenomas (≥1 cm), we used the estimate for advanced adenomas from Imperiale et al.,[14] who defined advanced adenomas as those with high-grade dysplasia or ≥25% villous histologic features or measuring ≥1 cm in size. As colonoscopy sensitivity only increases with size of the adenoma, the follow-up colonoscopy after a positive mtSDNA test will detect more high-grade dysplasia and adenomas with villous components if they are all assumed to be of large size, as opposed to if we had modeled histology and grade explicitly. Therefore, we may have overestimated the effectiveness of mtSDNA testing.

Third, the simulated cohort did not have any CRC screening prior to age 65. In practice, the Medicare population increasingly exists of individuals who already had some type of CRC screening. Although the effectiveness of screening is lower for these individuals, the relative difference in effectiveness and cost-effectiveness between different screening strategies is expected to be similar.

Fourth, we assumed conditional independence of repeat screenings. Consequently we assumed that there were no systematic false-negative results for adenomas and cancers. This assumption may not hold for gFOBT and FIT testing because bleeding of a lesion may not be a random event.[40] However, this may also not hold for mtSDNA testing because testing for blood is an important component of the test. Furthermore, the lesion in question may have acquired a gene mutation not assessed by the mtSDNA test, which means our assumption of conditional independence may be less likely to hold for the mtSDNA test compared to gFOBT and FIT. As a result, we may have overestimated the benefit of mtSDNA testing compared to the other tests, and its threshold reimbursement rate may be even lower than estimated here.

Finally, because test-specific data on longitudinal screening patterns are lacking, our base-case analysis assumes 100% adherence with screening, follow-up and surveillance procedures. Uptake of screening among the Medicare population is considerably less than 100%,[28] as is adherence with repeat screening[41], follow-up[42] and surveillance[43]. Meanwhile, overuse of resources is also common,[44] and a positive stool test is sometimes followed by another stool test instead of by the prescribed follow-up colonoscopy.[45] Although we did include a sensitivity analysis with 69% uptake, we assumed that individuals were either fully adherent or fully non-adherent with screening. The impact of less-than-perfect adherence among those who take up screening will vary according to the interval of testing and the characteristics of the test.

In summary, our analysis shows that compared with no screening, triennial mtSDNA testing reduces CRC incidence and mortality. However, it is less effective than other CRC screening options available to Medicare beneficiaries. At its current reimbursement rate, triennial mtSDNA testing also has higher costs than all other strategies, making it an inefficient screening option. It could be efficient and potentially cost-effective if mtSDNA testing would increase the adherence with CRC screening to nearly 100%. Triennial (or more frequent) mtSDNA testing could also be potentially cost-effective if the reimbursement rate were substantially lower, i.e. similar to that of other stool-based tests.

REFERENCES

1. American Cancer Society. Cancer Facts & Figures 2016. Atlanta: American Cancer Society, 2016.
2. Mandel JS, Church TR, Bond JH, et al. The effect of fecal occult-blood screening on the incidence of colorectal cancer. N Engl J Med 2000;343:1603-7.
3. Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. Lancet 2010;375:1624-33.
4. Segnan N, Armaroli P, Bonelli L, et al. Once-only sigmoidoscopy in colorectal cancer screening: follow-up findings of the Italian Randomized Controlled Trial--SCORE. J Natl Cancer Inst 2011;103:1310-22.
5. Schoen RE, Pinsky PF, Weissfeld JL, et al. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. N Engl J Med 2012;366:2345-57.
6. Holme O, Loberg M, Kalager M, et al. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: a randomized clinical trial. JAMA 2014;312:606-15.
7. Mandel JS, Church TR, Ederer F, et al. Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. J Natl Cancer Inst 1999;91:434-7.
8. Shaukat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for colorectal cancer. N Engl J Med 2013;369:1106-14.
9. Scholefield JH, Moss SM, Mangham CM, et al. Nottingham trial of faecal occult blood testing for colorectal cancer: a 20-year follow-up. Gut 2012;61:1036-40.
10. Kronborg O, Jorgensen OD, Fenger C, et al. Randomized study of biennial screening with a faecal occult blood test: results after nine screening rounds. Scand J Gastroenterol 2004;39:846-51.
11. Lindholm E, Brevinge H, Haglind E. Survival benefit in a randomized clinical trial of faecal occult blood screening for colorectal cancer. Br J Surg 2008;95:1029-36.
12. Faivre J, Dancourt V, Lejeune C, et al. Reduction in colorectal cancer mortality by fecal occult blood screening in a French controlled study. Gastroenterology 2004;126:1674-80.
13. U.S. Preventive Services Task Force, Bibbins-Domingo K, Grossman DC, et al. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. JAMA 2016;315:2564-75.
14. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med 2014;370:1287-97.
15. Centers for Medicare and Medicaid Services. Decision Memo for Screening for Colorectal Cancer - Stool DNA Testing (CAG-00440N), 2014.
16. Morson B. President's address. The polyp-cancer sequence in the large bowel. Proc R Soc Med 1974;67:451-7.
17. Morson BC. Evolution of cancer of the colon and rectum. Cancer 1974;34:suppl:845-9.
18. Morson BC. The evolution of colorectal carcinoma. Clin Radiol 1984;35:425-31.
19. American Cancer Society. Colorectal Cancer Screening: Insurance Coverage. Volume 2016: American Cancer Society, 2016.
20. Lieberman DA, Rex DK, Winawer SJ, et al. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 2012;143:844-57.
21. van Hees F, Zauber AG, Klabunde CN, et al. The appropriateness of more intensive colonoscopy screening than recommended in Medicare beneficiaries: a modeling study. JAMA Intern Med 2014;174:1568-76.
22. Warren JL, Klabunde CN, Mariotto AB, et al. Adverse events after outpatient colonoscopy in the Medicare population. Ann Intern Med 2009;150:849-57, W152.
23. Yabroff KR, Lamont EB, Mariotto A, et al. Cost of care for elderly cancer patients in the United States. J Natl Cancer Inst 2008;100:630-41.

24. U.S. Food and Drug Administration. FDA Approves Erbitux for Colrectal Cancer, 2004.
25. U.S. Food and Drug Administration. FDA Approves First Angiogenesis Inhibitor to Treat Colorectal Cancer, 2004.
26. Weinstein MC, Siegel JE, Gold MR, et al. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. JAMA 1996;276:1253-8.
27. Grosse SD. Assessing cost-effectiveness in healthcare: history of the $50,000 per QALY threshold. Expert Rev Pharmacoecon Outcomes Res 2008;8:165-78.
28. Sabatino SA, White MC, Thompson TD, et al. Cancer screening test use - United States, 2013. MMWR Morb Mortal Wkly Rep 2015;64:464-8.
29. Jones RM, Woolf SH, Cunningham TD, et al. The relative importance of patient-reported barriers to colorectal cancer screening. Am J Prev Med 2010;38:499-507.
30. Code of Federal Regulations (CFR) 414.508 - Payment for a new clinical diagnostic laboratory test. In: Government F, ed. 42, 2011.
31. Centers for Medicare and Medicaid Services (CMS). Clinical Laboratory Fee Schedule (CLFS): Final Determinations, 2015.
32. Protecting Access to Medicare Act of 2014. H.R. 4302 (113th), 2014.
33. Centers for Medicare and Medicaid Services (CMS). Medicare Program: Medicare Clinical Diagnostic Laboratory Tests Payment System (CMS-1621-F), 2016.
34. Zauber AG, Lansdorp-Vogelaar I, Wilschut J, et al. Cost-Effectiveness of DNA Stool Testing to Screen for Colorectal Cancer: Report to AHRQ and CMS from the Cancer Intervention and Surveillance Modeling Network (CISNET) for MISCAN and SimCRC Models. 2007. Available at: www.cms.hhs.gov/mcd/viewtechassess.asp?from2=viewtechassess.asp&id=212&.
35. Lansdorp-Vogelaar I, Kuntz KM, Knudsen AB, et al. Stool DNA testing to screen for colorectal cancer in the Medicare population: a cost-effectiveness analysis. Ann Intern Med 2010;153:368-77.
36. Ladabaum U, Mannalithara A. Comparative Effectiveness and Cost-Effectiveness of a Multi-target Stool DNA Test to Screen for Colorectal Neoplasia. Gastroenterology 2016.
37. Heigh RI, Yab TC, Taylor WR, et al. Detection of colorectal serrated polyps by stool DNA testing: comparison with fecal immunochemical testing for occult blood (FIT). PLoS One 2014;9:e85659.
38. Holme O, Bretthauer M, Eide TJ, et al. Long-term risk of colorectal cancer in individuals with serrated polyps. Gut 2015;64:929-36.
39. O'Brien MJ, Winawer SJ, Zauber AG, et al. The National Polyp Study. Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas. Gastroenterology 1990;98:371-9.
40. van der Meulen MP, Lansdorp-Vogelaar I, van Heijningen EM, et al. Nonbleeding adenomas: Evidence of systematic false-negative fecal immunochemical test results and their implications for screening effectiveness-A modeling study. Cancer 2016;122:1680-8.
41. Gellad ZF, Stechuchak KM, Fisher DA, et al. Longitudinal adherence to fecal occult blood testing impacts colorectal cancer screening quality. Am J Gastroenterol 2011;106:1125-34.
42. Partin MR, Burgess DJ, Burgess JF, Jr., et al. Organizational predictors of colonoscopy follow-up for positive fecal occult blood test results: an observational study. Cancer Epidemiol Biomarkers Prev 2015;24:422-34.
43. Schoen RE, Pinsky PF, Weissfeld JL, et al. Utilization of surveillance colonoscopy in community practice. Gastroenterology 2010;138:73-81.
44. Holden DJ, Jonas DE, Porterfield DS, et al. Systematic review: enhancing the use and quality of colorectal cancer screening. Ann Intern Med 2010;152:668-76.
45. Miglioretti DL, Rutter CM, Bradford SC, et al. Improvement in the diagnostic evaluation of a positive fecal occult blood test in an integrated health care organization. Med Care 2008;46:S91-6.

46. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, van Ballegooijen M, Kuntz KM. Evaluating test strategies for colorectal cancer screening: a decision analysis for the U.S. Preventive Services Task Force. Ann Intern Med. 2008;149(9):659-69.
47. Levi Z, Birkenfeld S, Vilkin A, Bar-Chana M, Lifshitz I, Chared M, et al. A higher detection rate for colorectal cancer and advanced adenomatous polyp for screening with immunochemical fecal occult blood test than guaiac fecal occult blood test, despite lower compliance rate. A prospective, controlled, feasibility study. Int J Cancer. 2011;128(10):2415-24.
48. Allison JE, Tekawa IS, Ransom LJ, Adrain AL. A comparison of fecal occult-blood tests for colorectal-cancer screening. N Engl J Med. 1996;334(3):155-9.
49. Schroy PC, 3rd, Coe A, Chen CA, O'Brien MJ, Heeren TC. Prevalence of advanced colorectal neoplasia in white and black patients undergoing screening colonoscopy in a safety-net hospital. Ann Intern Med. 2013;159(1):13-20.
50. van Rijn JC, Reitsma JB, Stoker J, Bossuyt PM, van Deventer SJ, Dekker E. Polyp miss rate determined by tandem colonoscopy: a systematic review. Am J Gastroenterol. 2006;101(2):343-50.
51. Zauber AG, Levin TR, Jaffe CC, Galen BA, Ransohoff DF, Brown ML. Implications of new colorectal cancer screening technologies for primary care practice. Med Care. 2008;46(9 Suppl 1):S138-46.
52. Johnson CD, Chen MH, Toledano AY, Heiken JP, Dachman A, Kuo MD, et al. Accuracy of CT colonography for detection of large adenomas and cancers. N Engl J Med. 2008;359(12):1207-17.
53. Weissfeld JL, Schoen RE, Pinsky PF, Bresalier RS, Church T, Yurgalevitch S, et al. Flexible sigmoidoscopy in the PLCO cancer screening trial: results from the baseline screening examination of a randomized trial. J Natl Cancer Inst. 2005;97(13):989-97.
54. Atkin WS, Cook CF, Cuzick J, Edwards R, Northover JM, Wardle J, et al. Single flexible sigmoidoscopy screening to prevent colorectal cancer: baseline findings of a UK multicentre randomised trial. Lancet. 2002;359(9314):1291-300.
55. Painter J, Saunders DB, Bell GD, Williams CB, Pitt R, Bladen J. Depth of insertion at flexible sigmoidoscopy: implications for colorectal cancer screening and instrument design. Endoscopy. 1999;31(3):227-31.

APPENDIX: CALCULATION OF SCREENING COSTS

Since the implementation of the Patient Protection and Affordable Care Act,[1] reimbursement rules have changed. As of January 2011, if a procedure was performed for screening and no biopsies or polypectomies were performed, a screening code (i.e., no biopsy or polypectomy) is used and patient coinsurance is waived (i.e., CMS reimburses 100% of the cost). If the procedure is diagnostic (i.e., follow-up exam after a positive stool-based test, regardless of findings) or therapeutic (i.e., biopsy or polypectomy is performed, regardless of whether it is a follow-up exam or originally for screening purposes), CMS covers 80% of the cost and the beneficiary is responsible for the remaining 20%. Because our analysis is from the perspective of CMS, these reimbursement rules are factored into all of our cost estimates.

Payments for the stool-based tests (i.e., gFOBT, FIT, and mtSDNA) were based on the 2014 Clinical Laboratory Fee Schedule[2] for gFOBT and FIT and the 2015 Fee Schedule[3] for mtSDNA, which was not a covered test in 2014. Average payments for endoscopic procedures (i.e., sigmoidoscopy and colonoscopy) were calculated from data provided by CMS and were based on 2014 outpatient Medicare claims data from the Chronic Conditions Data Warehouse (CCW).[4] For this analysis, three places of service were considered: physician office setting, outpatient prospective payment system (OPPS), and ambulatory surgical center (ASC). We excluded claims for inpatient endoscopic procedures because screening, follow-up, and surveillance endoscopies are not typically performed in that setting. For procedures performed in the OPPS or ASC setting, we included associated facility charges.

Estimated costs also include those associated with anesthesia services, anesthetic agents, and pathology when provided in conjunction with an endoscopy. Data on the frequency of use of anesthesia services during endoscopic procedures was obtained from 2013 Medicare claims data (personal communication, Leslie Narramore, CPC, MPA, of the American Gastroenterological Association). For colonoscopy, use of anesthesia services ranged from 54% for screening colonoscopies without lesion removal to 63% for colonoscopies performed for follow-up of positive findings on another screening test. Anesthesia services were used for 22% of screening sigmoidoscopies (**Appendix Table 1**). The cost of anesthesia services was derived using 2014 claims data from the CCW,[4] provided by CMS. Estimates ranged from $90-104 per procedure. We also included the cost of anesthesia medication (i.e., propofol), estimated at $3.75 per colonoscopy. This estimate was based on an assumption of 300 mg of propofol per procedure[5,6] and a payment limit of $0.125 per 10 mg[7] ($0.125 * 300 mg / 10 mg = $3.75).

Pathology costs are incurred when a lesion is found and removed during colonoscopy. Specimens are sent to the laboratory for pathology review in jars that typically include biopsy material from one segment of the bowel (i.e., if more than one biopsy is obtained from the descending colon, all of the specimens from that segment may be included in a single jar). If specimens are obtained from multiple segments of the bowel, specimens from each segment are typically placed in a separate jar. Each jar requires a separate pathology

service code. We estimated the mean number of jars per patient using data from the National Colonoscopy Study (personal communication, Ann G. Zauber, PhD), finding that the average number of jars per patient is approximately 1.4. The mean payment for pathological evaluation of one specimen jar was derived from 2014 claims data from the CCW[4] and estimated at $54.02. Combining this information, we estimated the average per-patient cost of pathology (when performed) as $75.63 (i.e., 1.4 * $54.02). We then applied this cost to each colonoscopy that is simulated to undergo pathology (i.e., any adenoma or colorectal cancer detected or a false positive colonoscopy). We assumed that no polypectomy was performed with flexible sigmoidoscopy screening.

If multiple lesions are detected within a single colonoscopy and all are biopsied or removed using the same technique, payment for the colonoscopy remains the same as it would be if only one lesion was intervened upon. However, if different types of polypectomy are required for the removal of multiple polyps, CMS reimburses 100% for the most expensive procedure code, and for each additional procedure code it reimburses the difference between that procedure and the base endoscopy code (i.e., the code for basic washing of the colon (CPT 45378)).[8] These reimbursement rules were captured in the data from the CCW provided by CMS that reported total payments by CMS for unique colonoscopies in which one or more surgical codes were submitted (i.e., CPT codes 45380-45381, 45383-45385. See **Appendix Table 2** for code descriptions). Estimates of the average cost per screening test based on these assumptions are provided in **Table 2**.

References

1. Patient Protection and Affordable Care Act, Pub. L. No. 111–148, 42 U.S.C. §§ 18001-18121, 2010.
2. Centers for Medicare and Medicaid Services. 2014 Clinical Laboratory Fee Schedule (CLAB) Public Use File (PUF). Available at www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ClinicalLabFeeSched/clinlab.html or www.cms.gov/apps/ama/license.asp?file=/ClinicalLabFeeSched/downloads/14CLAB.zip, 2014.
3. Centers for Medicare and Medicaid Services. 2015 Clinical Laboratory Fee Schedule (CLAB) Public Use File (PUF). Available at www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ClinicalLabFeeSched/clinlab.html or www.cms.gov/apps/ama/license.asp?file=/ClinicalLabFeeSched/Downloads/15CLAB.zip, 2015.
4. Chronic Condition Data Warehouse. Volume 2016.
5. Byrne MF, Chiba N, Singh H, et al. Propofol use for sedation during endoscopy in adults: a Canadian Association of Gastroenterology position statement. Can J Gastroenterol 2008;22:457-9.
6. Ramsay MA, Newman KB, Jacobson RM, et al. Sedation levels during propofol administration for outpatient colonoscopies. Proceedings 2014;27:12-5.
7. Centers for Medicare and Medicaid Services. Payment allowance limits for Medicare Part B drugs. , 2015.
8. Centers for Medicare and Medicaid Services. MLN Matters: Payment for Multiple Surgeries in a Method II Critical Access Hospital (CAH), 2012.

Appendix Table 1. Current Procedural Terminology codes and Healthcare Common Procedure Coding System codes used for colorectal cancer screening, follow-up, and surveillance tests and procedures, and mean reimbursement for the test/procedure and accompanying pathology and anesthesia services.

Test/procedure	CPT/HCPCS code	CPT/HCPCS description
gFOBT	82270	Guaiac-based fecal occult blood test
FIT	G0328	Immunochemical fecal occult blood test
mtSDNA	G0464	Stool-based DNA and immunochemical fecal occult blood test
Sigmoidoscopy[e]	G0104	Screening sigmoidoscopy
Screening COL without lesion removal	G0121	Screening colonoscopy; average risk
Follow-up COL (after positive finding on another screening test) without lesion removal	45378	Diagnostic colonoscopy for persons with signs/symptoms
Surveillance COL without lesion removal	G0105	Screening colonoscopy; high-risk
Any COL with lesion removal[f]	One or more of 45380-45381, 45383-45385	Colonoscopy with intervention (hot/cold biopsy, snare biopsy, other)

COL = colonoscopy; CPT code = Current Procedural Terminology code; FIT = fecal immunochemical test; gFOBT = guaiac-based fecal occult blood test; HCPCS code = Healthcare Common Procedure Coding System code; mtSDNA = multitarget stool DNA test.

[a] Reimbursement for pathology services was assumed to apply only to colonoscopy procedures in which polypectomy was performed.

[b] Frequency of claims for same-day anesthesia services were obtained from an analysis of 2013 Medicare claims (personal communication, Leslie Narramore of the American College of Gastroenterology).

[c] Includes facility payments, when appropriate.

[d] Weighted average payment across procedures with ($90-104) and without ($0) claims for anesthesia services. Payments for procedures with anesthesia services includes payment for the anesthetic agent, assuming reimbursement at $3.75 per procedure.

[e] Sigmoidoscopy is simulated without biopsy or polypectomy of detected lesions.

[f] If multiple polyps are removed by different methods (**Appendix Table 2**), more than one CPT code may be submitted for one colonoscopy. In such cases, payment for all but the highest-reimbursed procedure is reduced to the difference between the payment for the procedure of interest and the payment for basic washing of the colon (CPT 45378).

[g] Based on data from the National Colonoscopy Study, we assume there are on average 1.4 jars sent to pathology for every colonoscopy with polypectomy (1.4 * $54.02 = $75.63).

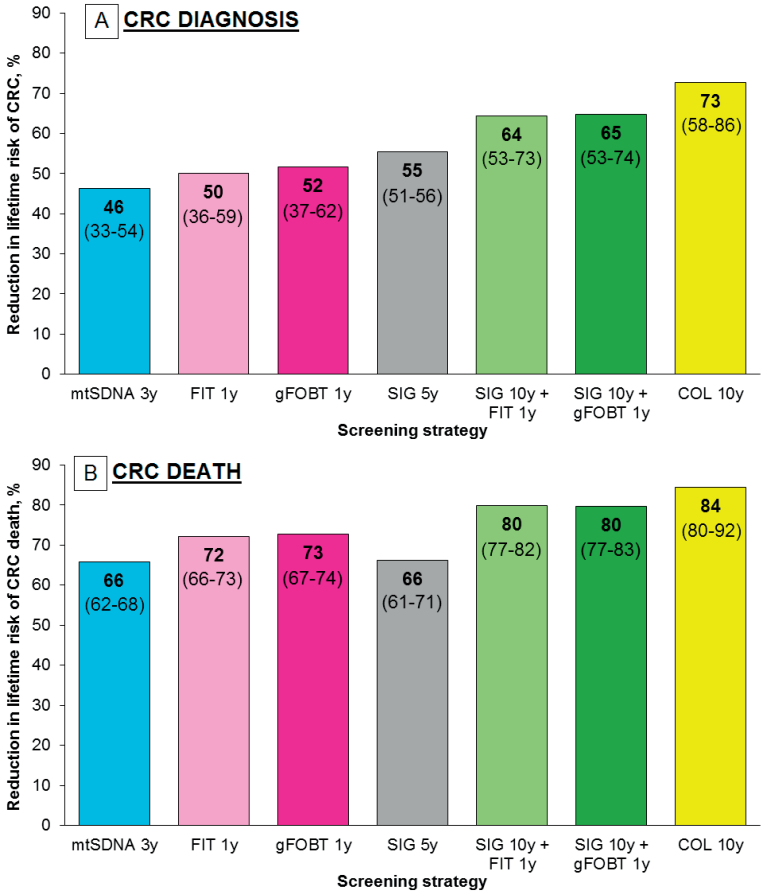
Frequency of accompanying claims (% of procedures with claim)		Mean payment, by component[c]			
Pathology services[a]	Anesthesia services[b]	Test/ procedure	Pathology payment	Anesthesia services[d]	Total
0	0	$4.44	$0	$0	$4.44
0	0	$21.70	$0	$0	$21.70
0	0	$492.72	$0	$0	$492.72
0	22	$253.90	$0	$20.57	$274.47
0	54	$647.03	$0	$51.68	$698.71
0	63	$524.40	$0	$66.47	$590.87
0	57	$626.38	$0	$55.01	$681.39
100[g]	61	$672.34	$75.63	$65.33	$813.31

Appendix Table 2. Descriptions of the surgical Current Procedural Terminology codes used for colonoscopies with biopsy or polypectomy.

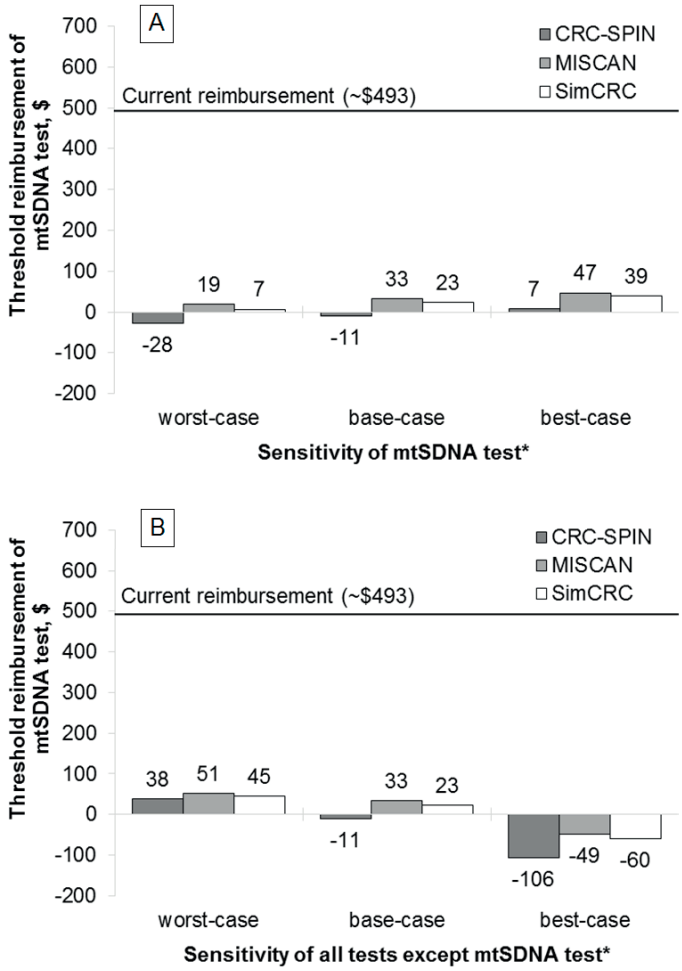
CPT code	Description
45380	Colonoscopy with biopsy, single or multiple (forceps to grab tissue w/o cautery)
45381	Colonoscopy with submucosal injection
45384	Colonoscopy with removal of tumor(s), polyp(s), or other lesion(s) by hot biopsy forceps
45385	Colonoscopy with removal of tumor(s), polyp(s), or other lesion(s) by snare technique
45383[a]	Colonoscopy with ablation of tumor(s), polyp(s) or other lesion(s) not amenable to removal by hot biopsy forceps, bipolar cautery or snare technique

CPT code = Current Procedural Terminology code

[a] Code was deleted in 2015 and replaced with 45388



Appendix Figure 1. Reductions in lifetime risk of being diagnosed with (Panel A) and dying from (Panel B) colorectal cancer. Median reduction and range across models are shown for each screening strategy.
COL = colonoscopy; CRC = colorectal cancer; FIT = fecal immunochemical test; gFOBT = high sensitivity guaiac-based fecal occult blood test; mtSDNA = multitarget stool DNA test; SIG = flexible sigmoidoscopy.



Appendix Figure 2. Sensitivity analyses: Reimbursement thresholds for the mtSDNA test at which the mtSDNA test strategy is efficient compared with other reimbursed CRC screening strategies for different levels of the sensitivity of the mtSDNA test (Panel A), and for different levels of the sensitivity of all other tests (Panel B).

* See Table 1 for the worst-case, base-case, and best-case test sensitivities for mtSDNA and all other tests.

PART IV

Reducing Harms and Increasing Benefits: Opportunities for Risk-based Screening

Chapter 8

Cervical cancer screening in partly HPV vaccinated cohorts – a cost-effectiveness analysis

Steffie K. Naber, Suzette M. Matthijsse, Kirsten Rozemeijer, Corine Penning, Inge M.C.M. de Kok, Marjolein van Ballegooijen

PLoS ONE. 2016 Jan; 11(1): e0145548.

ABSTRACT

Background: Vaccination against the oncogenic human papillomavirus (HPV) types 16 and 18 will reduce the prevalence of these types, thereby also reducing cervical cancer risk in unvaccinated women. This (measurable) herd effect will be limited at first, but is expected to increase over time. At a certain herd immunity level, tailoring screening to vaccination status may no longer be worth the additional effort. Moreover, uniform screening may be the only viable option. We therefore investigated at what level of herd immunity it is cost-effective to also reduce screening intensity in unvaccinated women.

Methods: We used the MISCAN-Cervix model to determine the optimal screening strategy for a pre-vaccination population and for vaccinated women (~80% decreased risk), assuming a willingness-to-pay of €50,000 per quality-adjusted life year gained. We considered HPV testing, cytology testing and co-testing and varied the start age of screening, the screening interval and the number of lifetime screens. We then calculated the incremental cost-effectiveness ratio (ICER) of screening unvaccinated women with the strategy optimized to the pre-vaccination population as compared to with the strategy optimized to vaccinated women, assuming different herd immunity levels.

Results: Primary HPV screening with cytology triage was the optimal strategy, with 8 lifetime screens for the pre-vaccination population and 3 for vaccinated women. The ICER of screening unvaccinated women 8 times instead of 3 was €28,085 in the absence of herd immunity. At around 50% herd immunity, the ICER reached €50,000.

Conclusion: From a herd immunity level of 50% onwards, screening intensity based on the pre-vaccination risk level becomes cost-ineffective for unvaccinated women. Reducing the screening intensity of uniform screening may then be considered.

INTRODUCTION

Infection with the human papillomavirus (HPV) has been identified as a necessary cause for cervical cancer.[1] Both the bivalent vaccine (targeting HPV-types 16/18), which is used in the Netherlands, and the quadrivalent vaccine (targeting HPV-types 6/11/16/18) are effective in preventing the two highly oncogenic types 16 and 18,[2,3] that are found in roughly 80% of invasive cervical cancers.[4] Recently, a nonavalent vaccine has been approved,[5] targeting seven oncogenic (and two non-oncogenic) HPV-types and thereby potentially preventing almost 90% of cervical cancers worldwide.[6]

In the Netherlands, a catch-up campaign targeted all 13- to 16-year-old girls in 2009. Since 2010, all 12-year-old girls are offered vaccination. The three-dose vaccination coverage has steadily increased from 49% in the 1993 birth cohort to 61% in the 2000 birth cohort.[7,8] In these partly vaccinated cohorts, the prevalence of HPV-16/18 infections is lower than in the pre-vaccination population. Therefore, unvaccinated women in those cohorts will be at lower risk for developing cervical cancer. While this indirect protective effect of vaccination, so-called herd immunity, will be limited at first, it is expected to increase over time.[9] It can be estimated by the percentage reduction in HPV-16/18 prevalence among unvaccinated women who were offered vaccination, as compared to totally unvaccinated cohorts. In the Netherlands, primary HPV screening will be implemented in 2016. From then, it could be relatively easy to monitor HPV-16/18 prevalence in unvaccinated women.

In many developed countries, vaccinated cohorts are approaching the start age of cervical cancer screening. Especially in settings where both vaccinated and unvaccinated women are well represented, it is unclear what screening strategy should be offered. In the youngest vaccinated cohorts (with limited herd immunity), vaccinated women are at much lower risk than unvaccinated women and screening based on vaccination status is likely more cost-effective than current uniform screening.[10-13] However, vaccinated women may not accept being offered less screening, solely because they adhered to vaccination guidelines. Screening based on vaccination status also requires the linkage of the screening invitational system with vaccination registries, which may not be (fully) possible in all settings.

As long as the follow-up of HPV vaccinated women in trials and population-based settings is not long enough to observe (statistical) differences in cervical cancer rates between vaccinated and unvaccinated cohorts, countries are reluctant to reduce the screening frequency. In the US, the same screening protocol is recommended for both vaccinated and unvaccinated women.[14,15] European guidelines even state that HPV vaccines cannot replace or modify current routine cervical cancer screening protocols.[16]

What is merely realized, is that women at reduced risk (due to either vaccination or herd immunity) could also be harmed by too intensive screening. These women will be offered more screening tests than needed, which increases their probability of being referred to the gynecologist in the absence of clinically relevant lesions. Women with abnormal cytology or HPV positive test results commonly experience fear, self-blame, distress and anxiety about cervical cancer, which reduces their quality of life.[17,18] The ethical justification of continuing

screening optimized to unvaccinated women instead of to those who adhered to vaccination guidelines, is therefore questionable. Moreover, it is probably very inefficient and cost-ineffective to do so. To avoid this inefficiency, screening should be optimized to vaccinated women as soon as unvaccinated women are substantially protected via herd immunity. We investigated at what level of herd immunity this would be justified for unvaccinated women.

METHODS

Using the MISCAN-Cervix model, we determined two optimal screening strategies: one for a pre-vaccination cohort, and one for a vaccinated cohort. To determine the level of herd immunity for which it would be cost-effective to replace the first strategy by the second, both strategies were applied to an unvaccinated cohort, assuming different levels of herd immunity.

MISCAN-Cervix model

The MISCAN-Cervix model, which is described in more detail in the **Model appendix**, was used to estimate costs and effects of different screening strategies.[19] In all of the analyses presented here, we simulated a cohort of 1 million women. While none of these women were assumed to be affected by vaccination when determining the optimal screening strategy for the pre-vaccination population, all of them were assumed to be vaccinated when determining the optimal screening strategy for vaccinated women. Both these optimal strategies were then applied to unvaccinated women assuming various herd immunity levels.

A fraction of these women will acquire HPV infections and/or develop cervical intraepithelial neoplasia (CIN) lesions. If these precursors progress to cervical cancer, women may die from the disease. If the population undergoes screening, the disease can be detected and treated in an earlier stage. As a result, cervical cancer death may be prevented or postponed.

The population at risk for cervical cancer was simulated based on demographic and hysterectomy data;[20,21] mortality from other causes was estimated using the observed age-specific mortality in the Netherlands in 2013.[20] The age-specific incidence of HPV infections that progress to cervical cancer was calibrated to the age-specific incidence of cervical cancer, which was obtained from the Netherlands Cancer Registry (NCR).[22] The age-specific incidence of pre-invasive lesions that do not progress to cervical cancer was calibrated so that the simulated detection rates of CIN lesions fit the observed detection rates in the Netherlands. These observed detection rates were obtained from the nationwide network and registry of histo- and cytopathology in the Netherlands (PALGA) for the period 2000-2007.[23] The incidence of high-risk HPV infections that do not progress to CIN was calibrated so that the simulated prevalence of all high-risk HPV infections fits the observed high-risk HPV prevalence.[24,25]

In the model, disease is subdivided into seven sequential stages: high-risk HPV infection, three pre-invasive stages (CIN 1, 2, and 3), and three invasive stages (International Federation of Gynecology and Obstetrics (FIGO) stages IA, IB, and II+).[26] Pre-invasive and FIGO IA stages can be diagnosed by screening only, because no symptoms will develop, whereas

stages IB and II+ can also be clinically diagnosed. As precursors are usually not progressive;[27] in the model, most HPV infections will clear without ever resulting in neoplasia, and lesions in pre-invasive stages can regress spontaneously. In the hypothetical situation without competing other-cause mortality, undetected preclinical invasive neoplasia will always progress to clinical cancer. CIN 1 and CIN 2 can develop in the absence of a high-risk HPV infection; in that case the lesion will always regress. CIN 3 or worse can only develop if a high-risk HPV infection is present.[28]

Screening policies

We simulated four different screening policies: (A) primary HPV screening with reflex cytology triage and cytology triage after six months (future Dutch screening program), (B) primary cytology with reflex HPV triage, (C) combined primary HPV and cytology (i.e. co-testing) with HPV triage after 12 months, and (D) primary cytology with cytology and HPV triage after six months and cytology triage after 18 months (current Dutch screening program). Policies (A) and (B) were already found to be cost-effective in case of no herd immunity;[29]policies (C) and (D) are included because of their resemblance with current practice in the US and in the Netherlands, respectively.

Screening schedules

Screening schedules differed by start age, screening interval and number of screens in a lifetime. Possible start ages were 25, 30, 35, 40, and 45 years. The screening interval varied from 5 to 20 years and the number of lifetime screens ranged from 1 to 12. Because screening women older than 80 years is not likely to be beneficial,[30] all strategies ended at or before the age of 80. In this way, 354 screening schedules were created.

Assumptions for screening and treatment

As we aimed to optimized screening for women who adhere to screening guidelines, we assumed full attendance in both primary screening and triage testing (**Appendix Table 1**). The sensitivity of cytology (the probability that the result is at least atypical squamous cells of undetermined significance (ASCUS)) was assumed to be 40% for CIN 1, 50% for CIN 2 and 75% for CIN 3 or cancer.[31] In the model calibration, the sensitivity of detecting at least high-grade squamous intraepithelial lesion (HSIL) was estimated to be 4% for CIN 1, 18% for CIN 2, 56% for CIN 3 and 60% for cervical cancer. The specificity of cytology was estimated at 97.6%. Based on the observed difference in CIN 3 or cancer detection rates between cytology and the HPV test, we assumed the sensitivity of the HPV test to be 85% for a high-risk HPV infection.[32] Although contamination and cross-reactivity may cause HPV tests to produce positive results in the absence of high-risk HPV infections, we assumed the specificity for the presence of HPV to be 100% and modeled a possible lack in specificity by including fast-clearing infections.

Detection of pre-invasive lesions and their associated management, including treatment if necessary, were assumed to lead to a 100% cure rate. A woman can, however, acquire new

HPV infections and develop CIN lesions after CIN treatment. For invasive cancer, we determined age-specific and stage-specific survival probabilities based on data from the NCR.[33] Since cancers detected by screening are found in an earlier stage than clinically diagnosed ones, women have a higher chance of survival. Using the NCR data, we estimated that if an invasive cancer is screen-detected, the probability to die from cervical cancer is reduced by 89.4%, 50% and 20% for FIGO stages IA, IB and II+, respectively.[33]

Assumptions for costs and utility losses

The estimated costs are based on a societal perspective, and are reported in 2013 euros **Appendix Table 2**). Screening costs include the costs for the invitational system and quality assurance, time and travel costs of the woman being screened, costs of smear taking, costs of evaluating the smear, costs of repeat tests after an inadequate test result, and costs of registration in PALGA. Diagnosis costs for women referred for colposcopy, treatment costs for detected pre-invasive lesions, treatment costs for invasive cervical cancer and costs of palliative care were derived from previous cost studies performed in the Netherlands.[34] A small (psychological) loss in quality of life was assumed for attending screening (including waiting for the result) and for being in triage (including attending follow-up screenings).[35] Larger losses in quality of life were assumed for being diagnosed and treated for CIN or cancer, and for having a terminal stage of cervical cancer.[36,37] Both costs and health effects were discounted with an annual rate of 3%.

Assumptions for vaccination

We assumed the efficacy of the bivalent vaccine as observed in the PATRICIA trial,[42,43] which is 25.3% for HPV infections without cytological abnormalities,[38] and 35.0%, 54.8%, and 93.2% for CIN 1, 2, and 3, respectively (**Table 1**).[2] As vaccination trials have not showed any waning in vaccine efficacy until now,[39] the protection from vaccination was assumed to be lifelong. Due to limited follow-up of the trials, a reduction in cervical cancer incidence has not been observed yet. However, studies do give estimates of the type-specific reduction in HPV prevalence.[40,41] In combination with the HPV-type distribution observed in cervical cancer cases in western Europe,[4] the vaccine efficacy for cervical cancer was estimated at 83.4%. In this calculation we assumed that all cervical cancers are caused by a single oncogenic HPV-type, thereby avoiding overestimating the effect of the vaccine. We further assumed that all oncogenic types are equally likely to be co-infected with other oncogenic types, and decreased all type-specific HPV-positivity rates with the same percentage (6.6%) to account for multiple infections.

In the absence of herd immunity, unvaccinated women were assumed to have the cervical cancer risk as is currently observed in the Netherlands.[42] Full herd immunity was assumed to be equally effective as vaccination in preventing both HPV infections, CIN lesions and cervical cancer. When the herd immunity level was assumed to be e.g. 25%, then 25% of the infections, lesions and cancers that would have been prevented by vaccination, were averted in unvaccinated women.

Table 1. Vaccination assumptions for base-case analysis and sensitivity analyses.

			Vaccine efficacy				
	Vaccine type	**Vaccine duration**[b]	HPV infections without CIN	CIN grade 1	CIN grade 2	CIN grade 3	Cervical cancer
Directly observed from PATRICIA trial (base case)	Bivalent	Lifelong	25.3%	35.0%	54.8%	93.2%	83.4%[c]
Directly observed from FUTURE trial	Quadrivalent	Lifelong	21.4%[d]	29.7%	42.9%	45.5%	77.8%[c]
Indirectly based on PATRICIA trial[a]	Bivalent	Lifelong	51.4%	33.5%	55.4%	62.2%	83.4%
Indirectly based on FUTURE trial[a]	Quadrivalent	Lifelong	38.2%	26.1%	47.5%	53.9%	77.8%

HPV = human papillomavirus; CIN = cervical intraepithelial neoplasia.

[a] Vaccine efficacy is calculated by combining the reduction in type-specific HPV infections observed in the trial, with the HPV-type distribution observed in HPV infections without cytological abnormalities (in the Netherlands),[43] and in CIN 1, 2, and 3, and cervical cancer (in western Europe).[4]

[b] Trials do not (yet) show that vaccine efficacy wanes; we assumed that if it would, vaccine boosters would be offered.

[c] Because the follow-up of the trials is too short to give (meaningful) estimates for cervical cancer, we used the estimates from the indirect approach.

[d] Observed vaccine efficacy for high-risk HPV infections combined with ASCUS (atypical squamous cells of undetermined significance), trial results do not include efficacy for high-risk HPV infections only.

Analyses and outcomes

For a pre-vaccination and a vaccinated cohort, we simulated the screening strategies described earlier and determined their discounted costs and effects as compared to no screening. For both cohorts, the optimal screening strategy was determined as follows. We first excluded all dominated screening strategies, i.e. those strategies that were more costly and less effective than (combinations of) other strategies. We then ranked the efficient strategies based on the number of quality-adjusted life years (QALYs) gained and calculated their incremental cost-effectiveness ratio (ICER), i.e. the additional costs per additional QALY gained compared to the next less effective, efficient strategy. For each cohort, the optimal screening strategy was then defined as the strategy with an ICER just below the willingness-to-pay threshold of €50,000 per QALY gained, which is a commonly used threshold in cost-effectiveness analyses for cervical cancer screening.[29,43]

The two optimal screening strategies were applied to unvaccinated women assuming herd immunity levels of 0%, 25%, 50%, 75%, and 100%. For all these levels, the ICER of screening optimized to the pre-vaccination cohort as compared to screening optimized to the vaccinated cohort was calculated. If the ICER reached above €50,000 per QALY gained, screening optimized to the pre-vaccination risk level was no longer considered cost-effective for unvaccinated women.

Sensitivity analyses

In the sensitivity analyses, we varied the following parameters.

Vaccine efficacy

1. First, we used the vaccine efficacy from two randomized efficacy trials in which the quadrivalent vaccine was used (FUTURE I and II).[44,45] The efficacy found in these trials was lower than for the bivalent vaccine, i.e. 29.7%, 42.9%, and 45.5% for CIN 1, 2, and 3, respectively.[46] Because in these trials HPV testing was only used when cytological abnormalities were observed, the reduction in HPV infections in women without cytological abnormalities is not known. Instead, we used the reduction in HPV-positive women with ASCUS, which was 21.4%.[46] Again, the efficacy for cervical cancer was estimated using the type-specific reduction in HPV prevalence[41,47] and the HPV-type distribution in cervical cancer,[4] which resulted in an estimate of 77.8%.
2. Second, we estimated the efficacy for all disease stages by using the type-specific reduction in HPV prevalence observed in the PATRICIA trial and the HPV-type distribution observed in the Netherlands (for HPV infections without cytological abnormalities)[48] and in western Europe (for CIN lesions and cervical cancer).[4] This resulted in an assumed vaccine efficacy of 51.4% for HPV infections, and of 33.5%, 55.4%, and 62.2% for CIN 1, 2, and 3, respectively. For cervical cancer, the efficacy remained at its base-case value of 83.4%.
3. Finally, this indirect approach of combining the type-specific reduction in HPV prevalence with the HPV-type distribution in HPV infections, CIN lesions and cervical cancer was also used to determine the vaccine efficacy for the quadrivalent vaccine. The assumed vaccine efficacy was 42.6% for HPV infections, 28.6%, 50.6%, and 57.7% for CIN 1, 2, and 3, respectively, and 80.2% for cervical cancer.

Background risk for cervical cancer in unvaccinated women

Instead of assuming an equal background risk for vaccinated and unvaccinated women, we included two sensitivity analyses in which the background risk in unvaccinated women was assumed 50% higher and 50% lower than in vaccinated women.

RESULTS

Base-case analysis

For a pre-vaccination cohort, 6-yearly primary HPV screening in the age range 30-72 years is most cost-effective (**Appendix Table 3**). This corresponds to 8 screens in a lifetime. The optimal strategy for vaccinated women is also primary HPV screening, but in a smaller age range (35-59 years) and with a longer interval (every 12 years), corresponding with 3 lifetime screens (**Appendix Table 4**).

Health effects

As compared to screening 3 times, screening 8 times reduces cervical cancer deaths with 161 per 100,000 unvaccinated women in the absence of herd immunity, and with 28 in case of full herd immunity (**Table 2**). It thereby yields 388 and 34 more QALYs gained when assuming 0% and 100% herd immunity, respectively (**Table 3**). However, it also requires more screen tests, more referrals for colposcopy and more CIN treatments. For one additionally prevented death, the required additional number of referrals for colposcopy increased from 34 for 0% herd immunity to 118 for 100%.

Table 2. Undiscounted health effects for unvaccinated women of primary HPV screening at ages 30-72 every 6 years (optimal for unvaccinated women without herd immunity) and at ages 35-59 every 12 years (optimal for vaccinated women), as compared to no screening. For different levels of herd immunity, results are given per 100,000 unvaccinated women.

Herd immunity level	Screening strategy	Primary screens	Triage screens	Referrals for colposcopy	False-positive referrals (no CIN)	CIN grade 1	CIN grade 2	CIN grade 3	Cases prevented	Deaths prevented
0%	30-72, 6y	717,049	55,427	10,188	873	3,805	2,360	3,029	1,416	589
	35-59, 12y	277,073	20,127	4,718	271	1,479	1,014	1,782	982	423
25%	30-72, 6y	716,804	51,324	8,969	823	3,630	2,080	2,340	1,123	471
	35-59, 12y	277,153	18,450	4,085	257	1,421	889	1,383	776	338
50%	30-72, 6y	716,579	47,130	7,756	770	3,468	1,802	1,648	832	348
	35-59, 12y	277,233	16,752	3,447	242	1,357	765	985	579	248
75%	30-72, 6y	716,354	42,929	6,535	723	3,286	1,528	953	537	225
	35-59, 12y	277,308	15,054	2,803	229	1,290	632	589	372	161
100%[a]	30-72, 6y	716,113	38,739	5,472	678	3,121	1,254	252	230	98
	35-59, 12y	277,386	13,352	2,156	213	1,228	511	176	158	70

CIN = cervical intraepithelial neoplasia.

[a] We assume that with full herd immunity, unvaccinated women have the same cervical cancer risk as vaccinated women.

Costs and cost-effectiveness

Screening 8 times instead of 3 increases total costs with approximately €10.9 and €11.1 million assuming no and full herd immunity, respectively. Consequently, the ICER of screening 8 times instead of 3 increased from €28,085 per QALY gained in the absence of herd immunity to €35,042, €47,530, €77,541, and €322,234 for 25%, 50%, 75%, and 100% herd immunity, respectively. From **Figure 1**, the estimated herd immunity level for which screening 8 times would cost approximately €50,000 per QALY gained when compared with screening 3 times, is 52%.

Table 3. Base-case costs and QALYs gained as compared to no screening (both 3% discounted) of screening optimized to a pre-vaccinated cohort and of screening optimized to a vaccinated cohort, and incremental cost-effectiveness of the former strategy as compared to the latter. For different levels of herd immunity, results are given per 100,000 unvaccinated women.

Herd immunity level	Screening strategy			Costs (€)	Incremental costs (€)	QALYs gained	Incremental QALYs	ICER (€)
	Age range	Interval	No. of screens					
0%	35-59	12y	3	4,458,721		1,488		
	30-72	6y	8	15,357,002	10,898,282	1,876	+388	28,085
25%	35-59	12y	3	5,062,986		1,184		
	30-72	6y	8	15,991,074	10,928,088	1,495	+312	35,042
50%	35-59	12y	3	5,756,793		868		
	30-72	6y	8	16,731,153	10,974,359	1,098	+231	47,530
75%	35-59	12y	3	6,457,603		556		
	30-72	6y	8	17,474,531	11,016,928	698	+142	77,541
100%[a]	35-59	12y	3	7,181,587		231		
	30-72	6y	8	18,277,092	11,095,505	265	+34	322,234

QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] With full herd immunity, unvaccinated women have the same cervical cancer risk as vaccinated women.

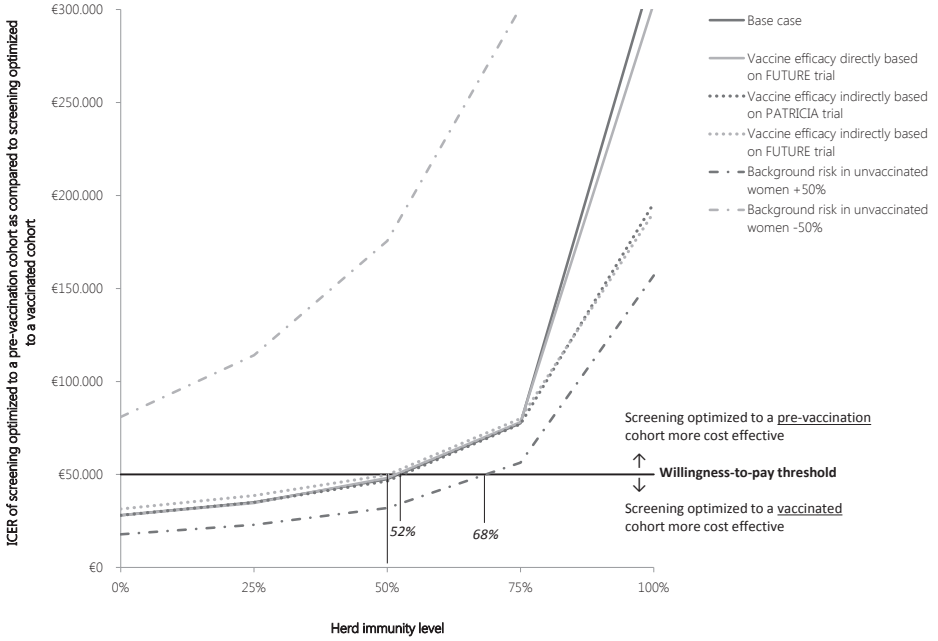


Figure 1. Incremental cost-effectiveness ratio (ICER) of screening optimized to a pre-vaccination cohort as compared to screening optimized to a vaccinated cohort, for unvaccinated women who benefit from different herd immunity levels, under both base-case assumptions and sensitivity analyses.

Sensitivity analyses

When vaccine efficacy was calculated indirectly from the FUTURE trial, the optimal screening strategy for vaccinated women involved an additional screening round at age 71 (**Appendix Table 5**). In all other sensitivity analyses, the optimal strategy for vaccinated women was unchanged (**Appendix Tables 6** and **7**).

Similar to the base-case analysis, the ICER of using the strategy optimized to the pre-vaccination cohort instead of to the vaccinated cohort, increased with increasing level of herd immunity (**Table 4**). In sensitivity analyses with different efficacy assumptions, screening optimized to the pre-vaccination population can be considered cost-effective as long as the herd immunity level is below 50-52%. When unvaccinated women would have a 50% lower background risk for cervical cancer, screening can be optimized to vaccinated women, regardless of the herd immunity level. If instead, unvaccinated women have a 50% higher background risk, screening optimized to the pre-vaccination population should be continued until the herd immunity reaches above ~68%.

Table 4. Results sensitivity analyses: Incremental cost-effectiveness ratio of screening optimized to a pre-vaccination cohort, as compared to screening optimized to a vaccinated cohort.

Herd immunity level	Vaccine efficacy[a]			Background risk in unvaccinated women	
	Directly observed from FUTURE trial	Indirectly based on PATRICIA trial	Indirectly based on FUTURE trial	+50%	-50%
0%	€28,085	€28,085	€31,450	€17,828	€80,972
25%	€35,050	€34,675	€38,631	€22,950	€114,122
50%	€46,471	€48,097	€49,747	€31,998	€175,596
75%	€77,153	€78,139	€80,122	€56,390	€301,129
100%[b]	€195,881	€303,352	€191,000	€157,043	QALYs lost[c]

QALYs = quality-adjusted life years.

[a] For vaccine efficacy assumptions, see **Table 1**.

[b] We assume that with full herd immunity, unvaccinated women have the same cervical cancer risk as vaccinated women.

[c] For unvaccinated women at 50% reduced cervical cancer risk, QALYs were lost when screening was optimized to the pre-vaccination risk level instead of to the risk level in vaccinated women.

DISCUSSION

For both a pre-vaccination and a vaccinated cohort, primary HPV screening is more cost-effective than primary cytology or co-testing. The optimal number of lifetime screens varied from 8 for the pre-vaccination cohort, to only 3 for the vaccinated cohort. For unvaccinated women, the adverse effects and costs of screening become more important as the herd immunity level increases. Offering these women 8 instead of 3 lifetime screens incremen-

tally required 34 colposcopy referrals per prevented death for 0% herd immunity, which increased to 118 referrals for 100% herd immunity. The ICER of screening 8 times instead of 3 increased from €28,085 per QALY gained in the absence of herd immunity to €322,234 at full herd immunity. Screening optimized to the risk level in vaccinated women becomes more cost-effective than screening optimized to the pre-vaccination risk level when the herd immunity reaches above 50-55%.

To foresee whether and when the herd immunity will reach this level, countries need to monitor the HPV-16/18 prevalence in unvaccinated women, starting with a reliable pre-vaccination baseline measurement. A recent cross-sectional study among women aged 18-24 years in Australia, in whom vaccination coverage was 55-74% for 1-3 doses,[49] showed a reduction in HPV-16/18 prevalence of 93% and 35% in vaccinated and unvaccinated women, respectively, compared to the pre-vaccination prevalence.[50] From these early data, the estimated herd immunity level would equal (0.35 / 0.93 ≈) 38%.

We have not incorporated vaccination coverage as a separate parameter in our analyses, the reason for which is as follows. Vaccination coverage plays a role in two ways: first, it determines how many unvaccinated women there are (which is important when evaluating how to screen them), and second, it is one of the main determinants of herd immunity. Mathematical models have been created to estimate the level of herd immunity given vaccination coverage.[51-53] These models have been helpful in decision analyses concerning vaccination (also in boys), by estimating its indirect effect in the unvaccinated. However, when it comes to screening decisions that depend on current or near future herd immunity, it seems more appropriate to seek guidance from actual measurements (of HPV prevalence in the unvaccinated) than from model based predictions of herd immunity levels. Indeed, the exact relation between coverage and herd immunity will only become established based on such measurements.

The manuscript primarily focused on the effect of decreasing the screening frequency of uniform screening for unvaccinated women. For vaccinated women, this adjustment would be cost-effective by definition. Meanwhile, it is important to point out that the harms of screening the vaccinated 8 times instead of 3 were smaller than the life years gained (**Table 3**), meaning that unadjusted screening did not result in a net loss in health for vaccinated women.

We optimized the screening strategy to the pre-vaccination risk level and to the risk level in vaccinated women. For partly vaccinated cohorts, it could be beneficial to have a screening strategy that is a compromise of these two strategies. In fact, when ignoring the costs and efforts related to restructuring screening guidelines, it would likely be cost-effective to reduce the screening frequency gradually while the herd immunity level increases. Adjusting national screening guidelines every few years is not a very workable solution though. Likewise, it could be cost-effective to tailor screening to vaccination status. Our results have shown that as soon as the herd immunity level reaches 50%, then it is beneficial (in terms of cost-effectiveness) for unvaccinated women to replace screening optimized to the pre-vaccination risk level with screening optimized to the risk level in vaccinated women. If this already happens within a few years, then establishing tailored screening by e.g. develop-

ing a vaccination registry that is linked to the screening invitational system, may not be worthwhile. The (lack of) accumulation of herd immunity over time is crucial in deciding whether the establishment of tailored screening would be worth these additional efforts. We performed our analyses under the assumption that it is most realistic that countries will continue screening all women uniformly, and that a once-only adjustment is made as soon as this seems justified for unvaccinated women.

Notable limitations are the following. First, we assumed that the efficacy of the vaccine has a lifelong duration. Although until now, HPV vaccination trials have shown a sustained efficacy,[2,3] it is possible that the efficacy will wane in the future. If the protection would fade away and offering vaccination boosters would not be an option, then screening optimized to vaccinated women would probably be more intensive than in the current analyses, and unvaccinated women could be screened accordingly from a lower herd immunity level onwards. Second, as the follow-up of the vaccination trials is too limited to give (meaningful) estimates of the vaccine efficacy for cervical cancer, we had to estimate this efficacy indirectly. The decrease in CIN 3 lesions does indicate that the vaccine is likely to prevent clinically relevant lesions, and therefore also cancer.[2,46] If the decrease in cervical cancer risk would be smaller than estimated, vaccinated women would also require more intensive screening, again meaning that unvaccinated women could be screened accordingly from a lower herd immunity level. Third, we assumed an equal background risk for vaccinated and unvaccinated women. Because reasons for refusing vaccination may vary widely (e.g. lack of knowledge about HPV, low perceived risk of infection, concerns about safety, religious values),[54] the background risk in unvaccinated women could both be higher or lower as compared to vaccinated women. In the sensitivity analyses we showed that even if the background risk in unvaccinated women would be 50% higher, then unvaccinated women could already be screened as vaccinated women from ~68% herd immunity onwards. Finally, we have not modeled the effects of the nonavalent vaccine, because its use is still limited compared to the bivalent and quadrivalent vaccine. If vaccination with this more potent vaccine would lead to a less intensive optimal screening strategy for vaccinated women, the herd immunity level at which unvaccinated women could be screened accordingly would be higher.

To our knowledge, this is the first study evaluating at what herd immunity level a once-only uniform (equal for vaccinated and unvaccinated women) screening adaptation becomes, considering risks, benefits and costs, an option. Because vaccinated women are approaching the age at which cervical cancer screening starts, the results of this study will be relevant in the near future. It shows, that as long as stepwise adjustment or dichotomized screening based on vaccination status are considered unfeasible, one may wait until the HPV-16/18 prevalence amongst unvaccinated women drops below 50% of the pre-vaccination level, before considering adjusting screening. Meanwhile, also the necessary evidence for a decrease in cervical cancer risk in vaccinated women should become available.

REFERENCES

1. Walboomers JM, Jacobs MV, Manos MM, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 1999;189:12-9.
2. Lehtinen M, Paavonen J, Wheeler CM, et al. Overall efficacy of HPV-16/18 AS04-adjuvanted vaccine against grade 3 or greater cervical intraepithelial neoplasia: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. Lancet Oncol 2012;13:89-99.
3. Future I/II Study Group, Dillner J, Kjaer SK, et al. Four year efficacy of prophylactic human papillomavirus quadrivalent vaccine against low grade cervical, vulvar, and vaginal intraepithelial neoplasia and anogenital warts: randomised controlled trial. BMJ 2010;341:c3493.
4. Guan P, Howell-Jones R, Li N, et al. Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer. Int J Cancer 2012;131:2349-59.
5. U.S. Food and Drug Administration. FDA approves Gardasil 9 for prevention of certain cancers caused by five additional types of HPV, Dec 10 2014.
6. Serrano B, Alemany L, Tous S, et al. Potential impact of a nine-valent vaccine in human papillomavirus related cervical disease. Infect Agent Cancer 2012;7:38.
7. van 't Schurink-van 't Klooster TM, Kemmeren JM, Vermeer-de Bondt PE, et al. Human papillomavirus vaccination catch-up campaign in 2009 for girls born in 1993 to 1996 in the Netherlands. Bilthoven: National Institute for Public Health and the Environment, 2011:53.
8. van Lier EA, Oomen PJ, Conyn-van Spaendonck MAE, et al. Immunisation coverage National Immunisation Programme in the Netherlands : Year of report 2015. Bilthoven: National Institute for Public Health and the Environment, 2015.
9. Drolet M, Benard E, Boily MC, et al. Population-level impact and herd effects following human papillomavirus vaccination programmes: a systematic review and meta-analysis. Lancet Infect Dis 2015;15:565-80.
10. Diaz M, de Sanjose S, Ortendahl J, et al. Cost-effectiveness of human papillomavirus vaccination and screening in Spain. Eur J Cancer 2010;46:2973-85.
11. Burger EA, Ortendahl JD, Sy S, et al. Cost-effectiveness of cervical cancer screening with primary human papillomavirus testing in Norway. Br J Cancer 2012;106:1571-8.
12. Goldhaber-Fiebert JD, Stout NK, Salomon JA, et al. Cost-effectiveness of cervical cancer screening with human papillomavirus DNA testing and HPV-16,18 vaccination. J Natl Cancer Inst 2008;100:308-20.
13. Coupe VM, Bogaards JA, Meijer CJ, et al. Impact of vaccine protection against multiple HPV types on the cost-effectiveness of cervical screening. Vaccine 2012;30:1813-22.
14. Saslow D, Solomon D, Lawson HW, et al. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. Am J Clin Pathol 2012;137:516-42.
15. Moyer VA, U.S. Preventive Services Task Force. Screening for cervical cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2012;156:880-91, W312.
16. European Centre for Disease Prevention and Control. Introduction of HPV vaccines in European Union countries - an update. In: ECDC, ed. Stockholm, 2012.
17. Herzog TJ, Wright JD. The impact of cervical cancer on quality of life--the components and means for management. Gynecol Oncol 2007;107:572-7.
18. Heinonen A, Tapper AM, Leminen A, et al. Health-related quality of life and perception of anxiety in women with abnormal cervical cytology referred for colposcopy: an observational study. Eur J Obstet Gynecol Reprod Biol 2013;169:387-91.

19. Habbema JD, van Oortmarssen GJ, Lubbe JT, et al. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed 1985;20:79-93.
20. Statline Database. The Hague: Statistics Netherlands (CBS), 2013.
21. Hospital Diagnosis Statistics 1963-1985. Utrecht: SIG (Information Centre for Health Care), 1985.
22. Netherlands Cancer Registry. Cervical cancer incidence by one-year age groups, 2000-2010., 2012.
23. Casparie M, Tiebosch AT, Burger G, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol 2007;29:19-24.
24. Bulkmans NW, Rozendaal L, Snijders PJ, et al. POBASCAM, a population-based randomized controlled trial for implementation of high-risk HPV testing in cervical screening: design, methods and baseline data of 44,102 women. Int J Cancer 2004;110:94-101.
25. Lenselink CH, Melchers WJ, Quint WG, et al. Sexual behaviour and HPV infections in 18 to 29 year old women in the pre-vaccine era in the Netherlands. PLoS One 2008;3:e3743.
26. Shepherd JH. Revised FIGO staging for gynaecological cancer. Br J Obstet Gynaecol 1989;96:889-92.
27. Nobbenhuis MA, Helmerhorst TJ, van den Brule AJ, et al. Cytological regression and clearance of high-risk human papillomavirus in women with an abnormal cervical smear. Lancet 2001;358:1782-3.
28. Stanley M. Pathology and epidemiology of HPV infection in females. Gynecol Oncol 2010;117:S5-10.
29. van Rosmalen J, de Kok IM, van Ballegooijen M. Cost-effectiveness of cervical cancer screening: cytology versus human papillomavirus DNA testing. BJOG 2012;119:699-709.
30. Castanon A, Landy R, Cuzick J, et al. Cervical screening at age 50-64 years and the risk of cervical cancer at age 65 years and older: population-based case control study. PLoS Med 2014;11:e1001585.
31. Berkhof J, de Bruijne MC, Zielinski GD, et al. Natural history and screening model for high-risk human papillomavirus infection, neoplasia and cervical cancer in the Netherlands. Int J Cancer 2005;115:268-75.
32. Rijkaart DC, Berkhof J, Rozendaal L, et al. Human papillomavirus testing for the detection of high-grade cervical intraepithelial neoplasia and cancer: final results of the POBASCAM randomised controlled trial. Lancet Oncol 2012;13:78-88.
33. Netherlands Cancer Registry (NCR). Age-specific and stage-specific cervical cancer incidence (2000-2010) and survival (1989-2009), 2012.
34. van Ballegooijen M, Rebolj M, Essink-Bot ML, et al. De effecten en kosten van het bevolkingsonderzoek naar baarmoederhalskanker in Nederland na de herstructurering. Rotterdam: Erasmus MC, afdeling Maatschappelijke Gezondheidszorg, 2006.
35. van Ballegooijen M. Effects and Costs of Cervical Cancer Screening. Department of Public Health. Rotterdam, the Netherlands: Erasmus University, 1998.
36. Mandelblatt JS, Lawrence WF, Womack SM, et al. Benefits and costs of using HPV testing to screen for cervical cancer. JAMA 2002;287:2372-81.
37. Goldie SJ, Kohli M, Grima D, et al. Projected clinical benefits and cost-effectiveness of a human papillomavirus 16/18 vaccine. J Natl Cancer Inst 2004;96:604-15.
38. Paavonen J, Naud P, Salmeron J, et al. Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women. Lancet 2009;374:301-14.
39. Schiller JT, Castellsague X, Garland SM. A review of clinical trials of human papillomavirus prophylactic vaccines. Vaccine 2012;30 Suppl 5:F123-38.
40. Wheeler CM, Castellsague X, Garland SM, et al. Cross-protective efficacy of HPV-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by non-vaccine oncogenic HPV types: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. Lancet Oncol 2012;13:100-10.

41. Malagon T, Drolet M, Boily MC, et al. Cross-protective efficacy of two human papillomavirus vaccines: a systematic review and meta-analysis. Lancet Infect Dis 2012;12:781-9.
42. Netherlands Cancer Registry (NCR). www.cijfersoverkanker, Incidence, survival, and mortality of cervical cancer, 2016.
43. Accetta G, Biggeri A, Carreras G, et al. Is human papillomavirus screening preferable to current policies in vaccinated and unvaccinated women? A cost-effectiveness analysis. J Med Screen 2010;17:181-9.
44. Garland SM, Hernandez-Avila M, Wheeler CM, et al. Quadrivalent vaccine against human papillomavirus to prevent anogenital diseases. N Engl J Med 2007;356:1928-43.
45. Future II Study Group. Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. N Engl J Med 2007;356:1915-27.
46. Munoz N, Kjaer SK, Sigurdsson K, et al. Impact of human papillomavirus (HPV)-6/11/16/18 vaccine on all HPV-associated genital diseases in young women. J Natl Cancer Inst 2010;102:325-39.
47. Brown DR, Kjaer SK, Sigurdsson K, et al. The impact of quadrivalent human papillomavirus (HPV; types 6, 11, 16, and 18) L1 virus-like particle vaccine on infection and disease due to oncogenic nonvaccine HPV types in generally HPV-naive women aged 16-26 years. J Infect Dis 2009;199:926-35.
48. Coupe VM, Berkhof J, Bulkmans NW, et al. Age-dependent prevalence of 14 high-risk HPV types in the Netherlands: implications for prophylactic vaccination and screening. Br J Cancer 2008;98:646-51.
49. Brotherton JM, Murray SL, Hall MA, et al. Human papillomavirus vaccine coverage among female Australian adolescents: success of the school-based approach. Med J Aust 2013;199:614-7.
50. Tabrizi SN, Brotherton JM, Kaldor JM, et al. Assessment of herd immunity and cross-protection after a human papillomavirus vaccination programme in Australia: a repeat cross-sectional study. Lancet Infect Dis 2014;14:958-66.
51. Brisson M, van de Velde N, Franco EL, et al. Incremental impact of adding boys to current human papillomavirus vaccination programs: role of herd immunity. J Infect Dis 2011;204:372-6.
52. Elbasha EH, Dasbach EJ, Insinga RP. Model for assessing human papillomavirus vaccination strategies. Emerg Infect Dis 2007;13:28-41.
53. Choi YH, Jit M, Gay N, et al. Transmission dynamic modelling of the impact of human papillomavirus vaccination in the United Kingdom. Vaccine 2010;28:4091-102.
54. Dempsey AF, Abraham LM, Dalton V, et al. Understanding the reasons why mothers do or do not have their adolescent daughters vaccinated against human papillomavirus. Ann Epidemiol 2009;19:531-8.

Appendix Table 1. Base-case assumptions for screening.

Parameter	Value
Attendance	100%
Cytology	
Probability of at least ASCUS for:	
CIN grade 1	40%
CIN grade 2	50%
CIN grade 3 or worse	75%
Probability of at least HSIL for:	
CIN grade 1	4%
CIN grade 2	18%
CIN grade 3	56%
Cervical cancer	60%
Specificity[a]	97.6%
HPV test	
Sensitivity for high-risk HPV infection	85%
Specificity for high-risk HPV infection	100%[b]

ASCUS = atypical squamous cells of undetermined significance; CIN = cervical intraepithelial neoplasia; HSIL = high-grade squamous intraepithelial lesion; HPV = human papillomavirus.

[a] Probability of a normal test result in women without CIN or cancer.

[b] Potential false-positive HPV test results were modeled as HPV infections with a short duration.

Appendix Table 2. Base-case assumptions for costs and utilities.

	Costs (€)	Utilities		
		Disutility	Duration	Quality-adjusted time lost
Invitation				
	4.91	-	-	-
Primary screening				
Cytology	66.95			
HPV test	63.63	0.005	2 weeks	2 hours
Cytology + HPV test	96.32			
Reflex triage				
Cytology	32.69	-	-	-
HPV test	29.38			
Triage after 6, 12 or 18 months				
Cytology	64.41	0.005	Time since last test	Depends on interval
HPV test	61.10	0.005		
Diagnosis and treatment of pre-invasive stages				
False-positive referral	300	0.005	0.5 year	22 hours
CIN grade 1	936	0.03	0.5 year	6 days
CIN grade 2	1,386	0.07	1 year	26 days
CIN grade 3	1,623	0.07	1 year	26 days
Diagnosis and treatment of cancer				
FIGO IA	5,314	0.062	5 years	4 months
FIGO IB	12,601	0.062	5 years	4 months
FIGO II+ (screen-detected)	12,420	0.28	5 years	17 months
FIGO II+ (clinically detected)	11,599	0.28	5 years	17 months
Terminal care				
	28,220	0.740	1 year	9 months

HPV = human papillomavirus; CIN = cervical intraepithelial neoplasia; FIGO = International Federation of Gynecology and Obstetrics.

Costs are in 2013 prices. €1.00 (£0.85; $1.37).

Appendix Table 3. Cost-effective strategies for a pre-vaccination cohort under base-case assumptions.

Strategy				Cost-effectiveness[a]		
Policy	**Age range**	**Interval**	**No. of screens**	**QALYs gained**	**Costs (€)**	**ICER (€)**
Primary HPV with cytology triage	45	-	1	817	3,656	-
Primary HPV with cytology triage	40 - 59	19	2	1,192	595,896	1,851
Primary HPV with cytology triage	40 - 57	17	2	1,210	654,022	3,093
Primary HPV with cytology triage	40 - 66	13	3	1,355	1,225,940	3,971
Primary HPV with cytology triage	35 - 71	12	4	1,545	2,777,862	8,138
Primary HPV with cytology triage	35 - 65	10	4	1,602	3,247,901	8,226
Primary HPV with cytology triage	35 - 71	9	5	1,662	4,013,836	12,866
Primary HPV with cytology triage	35 - 75	8	6	1,706	4,888,675	19,898
Primary HPV with cytology triage	35 - 70	7	6	1,726	5,505,034	30,264
Primary HPV with cytology triage	30 - 72	7	7	1,820	8,797,780	35,292
Primary HPV with cytology triage	**30 - 72**	**6**	**8**	**1,857**	**10,423,560**	**43,175**
Primary HPV with cytology triage	30 - 78	6	9	1,865	10,849,457	55,738
Primary HPV with cytology triage	30 - 75	5	10	1,890	13,024,210	85,673
Primary HPV with cytology triage	30 - 80	5	11	1,893	13,420,746	138,221
Primary HPV with cytology triage	30 - 74	4	12	1,897	16,495,407	865,290

QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio; HPV = human papillomavirus.

[a] Costs and QALYs were discounted at an annual rate of 3.

Appendix Table 4. Cost-effective strategies for a vaccinated cohort under base-case assumptions.

Strategy				Cost-effectiveness[a]		
Policy	**Age range**	**Interval**	**No. of screens**	**QALYs gained**	**Costs (€)**	**ICER (€)**
Primary HPV with cytology triage	45	-	1	115	1,618,507	-
Primary HPV with cytology triage	40	-	1	137	2,005,709	17,306
Primary HPV with cytology triage	40 - 57	17	2	178	3,057,518	25,930
Primary HPV with cytology triage	40 - 55	15	2	181	3,143,604	29,183
Primary HPV with cytology triage	35 - 50	15	2	198	3,855,345	39,937
Primary HPV with cytology triage	35 - 65	15	3	217	4,656,004	42,850
Primary HPV with cytology triage	**35 - 59**	**12**	**3**	**226**	**5,081,409**	**45,286**
Primary HPV with cytology triage	35 - 55	10	3	233	5,408,214	52,823
Primary HPV with cytology triage	35 - 65	10	4	242	6,238,514	88,735
Primary HPV with cytology triage	35 - 67	8	5	252	7,594,621	138,443
Primary HPV with cytology triage	35 - 75	8	6	254	8,139,809	257,058
Primary cytology with HPV triage	30 - 72	6	8	262	13,056,727	593,680
Primary cytology with HPV triage	30 - 78	6	9	263	13,534,997	862,056
Primary cytology with HPV triage	30 - 75	5	10	263	15,697,980	5,376,727

QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio; HPV = human papillomavirus.

[a] Costs and QALYs were discounted at an annual rate of 3%.

Appendix Table 5. Cost-effective strategies for a vaccinated cohort when vaccine efficacy is indirectly based on the FUTURE trial.

Strategy				Cost-effectiveness[a]		
Policy	**Age range**	**Interval**	**No. of screens**	**QALYs gained**	**Costs (€)**	**ICER (€)**
Primary HPV with cytology triage	40 - 53	13	2	123	3,475,978	-
Primary HPV with cytology triage	40 - 66	13	3	147	4,199,297	29,996
Primary HPV with cytology triage	35 - 59	12	3	176	5,354,578	40,512
Primary HPV with cytology triage	**35 - 71**	**12**	**4**	**189**	**5,955,460**	**47,493**
Primary HPV with cytology triage	35 - 75	10	4	197	6,465,406	57,658
Primary HPV with cytology triage	35 - 71	9	5	209	7,359,065	74,476
Primary HPV with cytology triage	35 - 75	8	6	216	8,312,103	149,656
Primary HPV with cytology triage	30 - 78	8	7	229	10,955,327	203,002
Primary HPV with cytology triage	30 - 72	6	8	234	13,466,715	457,539
Primary HPV with cytology triage	30 - 78	6	9	235	13,922,136	639,319
Primary cytology with HPV triage	30 - 75	5	10	236	16,060,457	1,680,746

QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio; HPV = human papillomavirus.

[a] Costs and QALYs were discounted at an annual rate of 3%.

Appendix Table 6. Cost-effective strategies for a vaccinated cohort when vaccine efficacy is directly observed from the FUTURE trial.

Strategy				Cost-effectiveness[a]		
Policy	**Age range**	**Interval**	**No. of screens**	**QALYs gained**	**Costs (€)**	**ICER (€)**
Primary HPV with cytology triage	45	-	1	178	1,448,046	-
Primary HPV with cytology triage	40	-	1	216	1,850,074	10,678
Primary HPV with cytology triage	40 - 58	18	2	273	2,776,309	16,067
Primary HPV with cytology triage	40 - 54	14	2	280	2,946,291	25,127
Primary HPV with cytology triage	35 - 65	15	3	334	4,455,547	27,861
Primary HPV with cytology triage	35 - 61	13	3	341	4,728,923	44,804
Primary HPV with cytology triage	**35 - 59**	**12**	**3**	**344**	**4,877,601**	**48,674**
Primary HPV with cytology triage	35 - 71	12	4	355	5,477,508	52,364
Primary HPV with cytology triage	35 - 65	10	4	364	5,984,606	53,712
Primary HPV with cytology triage	35 - 71	9	5	378	6,872,248	67,009
Primary HPV with cytology triage	35 - 75	8	6	384	7,816,679	156,284
Primary cytology with HPV triage	30 - 78	6	9	395	13,392,036	480,256
Primary cytology with HPV triage	30 - 75	5	10	395	15,537,380	37,371,197

QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio; HPV = human papillomavirus.
[a] Costs and QALYs were discounted at an annual rate of 3%.

Appendix Table 7. Cost-effective strategies for a vaccinated cohort when vaccine efficacy is indirectly based on the PATRICIA trial.

Strategy				Cost-effectiveness[a]		
Policy	**Age range**	**Interval**	**No. of screens**	**QALYs gained**	**Costs (€)**	**ICER (€)**
Primary HPV with cytology triage	45	-	1	129	1,515,666	-
Primary HPV with cytology triage	40	-	1	152	1,908,685	17,332
Primary HPV with cytology triage	40 - 57	17	2	196	2,904,752	22,669
Primary HPV with cytology triage	40 - 55	15	2	198	2,990,397	49,102
Primary HPV with cytology triage	**35 - 59**	**12**	**3**	**237**	**4,931,273**	**49,180**
Primary HPV with cytology triage	35 - 55	10	3	243	5,243,447	56,753
Primary HPV with cytology triage	35 - 65	10	4	255	6,030,039	66,648
Primary HPV with cytology triage	35 - 75	10	5	259	6,543,607	131,391
Primary HPV with cytology triage	35 - 75	8	6	264	7,863,497	259,251
Primary cytology with HPV triage	30 - 78	6	9	269	13,401,036	1,028,756

QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio; HPV = human papillomavirus.
[a] Costs and QALYs were discounted at an annual rate of 3%.

Chapter 9

Age-specific screening intervals for people with a family history of colorectal cancer: a cost-effectiveness analysis

Steffie K. Naber, Karen M. Kuntz, Nora B. Henrikson, Marc S. Williams, Ned Calonge, Katrina A. B. Goddard, Doris T. Zallen, Theodore G. Ganiats, Elizabeth M. Webber, A. Cecile J.W. Janssens, Marjolein van Ballegooijen, Ann G. Zauber, Iris Lansdorp-Vogelaar

Submitted.

ABSTRACT

Background: Although studies have shown that the relative risk (RR) of colorectal cancer (CRC) for people with a positive family history decreases with increasing age, none of the screening recommendations specifies less frequent screening with increasing age. The aim of this study is to determine whether such a refinement would be cost-effective.

Methods: From literature we estimated familial RR of developing CRC by number of affected first-degree relatives (FDRs) and age of the person with FDRs, as compared to average-risk individuals of the same age. Based on these RRs, the microsimulation MISCAN model estimated costs and effects of colonoscopy screening strategies, varying in age range and interval, to determine the most cost-effective strategy for each age group and number of FDRs.

Results: For people with one affected FDR, comprising 92% of those with a positive family history, 3-yearly screening from age 40 is most cost-effective. If no adenomas are found, the screening interval can gradually be extended to 5, 7, and 10 years at ages 45, 55, and 65, respectively. From a cost-effectiveness perspective, individuals with more affected FDRs preferably start screening earlier and at shorter intervals, but can also reduce its frequency if no abnormalities are found.

Conclusions: For individuals with a constant level of family history over time, it is cost-effective to gradually increase the screening interval after several subsequent negative colonoscopies. If no adenomas develop, it is unlikely that these individuals are affected by genetic predisposition, so continuing intensive colonoscopy screening would provide little or no additional health benefit and is clearly not cost-effective.

INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer in the US, with over 130,000 new diagnoses and about 50,000 deaths estimated in 2016.[1] Familial adenomatous polyposis (FAP), polyposis syndromes and hereditary non-polyposis colorectal cancer (HNPCC, also known as Lynch syndrome) account for approximately 5% of all CRC cases.[2] Apart from these well-defined inherited syndromes, people with relatives diagnosed with CRC are at an increased risk for the disease.[3] Whereas approximately 9% of the population aged 30 to 70 has a positive family history of CRC, 20% of all CRC cases occur in people with a positive family history.[4] The more affected first-degree relatives (FDRs) an individual has, the higher his or her absolute and relative risk for developing CRC.[5]

Current recommendations state that individuals at average risk for CRC undergo colonoscopy every 10 years starting at age 50.[6-8] People with a family history of CRC are recommended to start screening earlier and/or with a shorter interval.[7] Even though CRC risk increases rapidly with the number of affected FDRs[5] and screening people with ≥3 FDRs more often than every 5 years was shown to be cost-effective[9], current guidelines do not distinguish between having 2 or 3, or even more affected FDRs.

Furthermore, the effectiveness of screening in individuals with a positive family history might be improved by allowing the screening intensity to vary with age. An evidence-based review has shown that the excess CRC risk that is associated with a given family history of CRC decreases with age.[10] For example, having a positive family history at young age (e.g. <45 years) is rare,[10] and is associated with a much higher RR for CRC as compared to average-risk individuals of the same age, than having a positive family history at older ages. In addition to distinguishing individuals with 2 affected FDRs from those with 3 or ≥4, the decrease in familial RR by age justifies an investigation of whether screening guidelines for people with a positive family history could be improved.

METHODS

We used the Microsimulation Screening Analysis-Colon (MISCAN-Colon) model to quantify the effectiveness and costs of colonoscopy screening in individuals with a family history of CRC. More precisely, we determined the optimal (i.e. most cost-effective) colonoscopy screening schedule based on age of the individual at risk (i.e., 30-44, 45-49, 50-54, 55-59, 60-64, 65-69, and 70+ years) and his/her number of affected FDRs (i.e., 1, 2, 3, and ≥4), where an FDR is defined as one's parent, sibling or offspring. For each level of family history, the most cost-effective ages to begin and end screening, and the age-group specific screening interval were determined assuming a willingness-to-pay threshold of $100,000 per quality-adjusted life year (QALY).

MISCAN-Colon

MISCAN-Colon is a well-established microsimulation model for CRC developed at the Department of Public Health of the Erasmus University Medical Center (Rotterdam, the Netherlands). The model's structure, underlying assumptions, and calibration are described in the **Model appendix**. In brief, MISCAN-Colon simulates the life histories of a large population of persons from birth to death. As each simulated person ages, one or more adenomas may develop. These adenomas can progress from small (≤5 mm), to medium (6-9 mm), to large size (≥10 mm). Some adenomas can develop into preclinical cancer, which may progress through to cancer stages I to IV. During each stage CRC may be diagnosed because of symptoms. Survival after clinical diagnosis is determined by the stage at diagnosis, the localization of the cancer, and the person's age.[11]

Screening will alter some of the simulated life histories. Some cancers will be prevented by the detection and removal of adenomas; other cancers will be detected in an earlier stage with a more favorable survival. However, screening can also result in serious complications and overdiagnosis and overtreatment of CRC (i.e. the detection and treatment of cancers that would not have been diagnosed without screening). By comparing a simulation of life histories with screening to a simulation of the same life histories without screening, MISCAN-Colon quantifies the effectiveness of screening, as well as the associated costs.

MISCAN-Colon was calibrated to the age-, stage-, and localization-specific incidence of CRC as observed in the Surveillance, Epidemiology, and End Results (SEER) Program before the introduction of screening (i.e. between 1975 and 1979),[12] and the age-specific prevalence and the multiplicity distribution of adenomas as observed in autopsy studies.[13-22] The preclinical duration of CRC and the adenoma dwell-time were calibrated to the rates of interval and surveillance detected cancers observed in randomized controlled trials evaluating screening using guaiac fecal occult blood tests and a once-only sigmoidoscopy.[23-27] The model allows for a large variability in progression rates, enabling some adenomas to progress to CRC within a few years.

Modeling familial risk

An increased risk for developing CRC can be modeled through an increased risk of developing adenomas, or by assuming that adenomas are more likely to progress to CRC. Whereas an increased adenoma prevalence has been observed in individuals with a positive family history, there is limited evidence for accelerated adenoma progression.[10] We therefore modeled familial risk by assuming that individuals with affected FDRs develop more adenomas. Although the probability that a single adenoma progresses to CRC is not changed, the increased probability of developing adenomas does imply that, on average, CRC is diagnosed at a younger age.

Familial risk cohorts

We simulated a cohort of 10 million people for every combination of age group at risk (i.e., 30-44, 45-49, 50-54, 55-59, 60-64, 65-69, and 70+ years) and level of family history (i.e. having

0, 1, 2, 3, or 4 FDRs). These 35 cohorts differed in their estimated RR for developing CRC, which was based on two studies.[5,28] Lifetime RRs for people having 1, 2, 3, or ≥4 affected FDRs were obtained from Taylor et al.[5] and the development of RR by age of the person at risk for CRC was obtained from Fuchs et al.[28] The linkage of a population-based resource with a computerized genealogy (including >2.3 million individuals) to statewide cancer registry records, enabled Taylor et al.[5] to retrospectively compute lifetime RRs for developing CRC for different levels of family history. Fuchs et al.[28] were able to estimate age-specific RRs in a prospective study including ~120,000 individuals who self-reported CRC diagnoses in themselves and their families. Due to the size of this study, only RRs for individuals with ≥1 affected FDRs were reported.[28] By assuming that the age distribution of RR does not depend on the level of family history, we transformed the lifetime RRs from Taylor et al.[5] into familial age-specific RR values. To this end, the RR of every age group in Fuchs et al.[28] was divided by the overall RR presented in Fuchs et al.[28], which was then multiplied with the lifetime RR from Taylor et al.[5] Using this calculation, we found the familial age-specific RRs given in **Table 1**. Although we recognize that having a single affected FDR is much more common than having 2, 3, or ≥4 affected FDRs (91.7% of those with at least one affected FDR versus 7.3%, 0.8%, and 0.1%, respectively),[5] we simulated cohorts of equal size to ensure stability of model outcomes.

Table 1. Model inputs: Estimated relative risk of developing colorectal cancer for people with a positive family history by age, as compared to the average-risk population of the same age (middle section). Values were computed by multiplying the lifetime RR (right column) with the age-specific RR for individuals with ≥1 affected FDR (bottom row).

		AGE GROUP OF PERSON TO BE SCREENED							Lifetime RR[a]
		30-44	**45-49**	**50-54**	**55-59**	**60-64**	**65-69**	**70+**	
FAMILY HISTORY	**1 affected FDR**	5.49	4.12	3.00	2.00	1.60	1.36	1.08	2.04
	2 affected FDRs	8.66	6.49	4.73	3.16	2.52	2.15	1.70	3.22
	3 affected FDRs	12.74[b]	9.55	6.96	4.65	3.71	3.16	2.50	4.73
	≥4 affected FDRs	28.99[b]	21.72[b]	15.84[b]	10.58[b]	8.45	7.20	5.70	10.77[c]
Age-specific RR for individuals with ≥1 affected FDR[d]		2.69	2.02	1.47	0.98	0.78	0.67	0.53	

FDR = first-degree relative; RR = relative risk.

[a] Lifetime RRs are based on data reported in Table 1 from Taylor et al.[5] In that study, the weighted average of the lifetime RRs presented equals ~0.936. To ensure that the population has an average RR of 1, we divided the reported RRs by ~0.936.

[b] In a sensitivity analysis, relative risks were truncated at 10.

[c] For ≥4 affected FDR, we computed a weighted average of the relative risk associated with having 4 and ≥5 affected FDRs (7.74 and 19.86 respectively). We merged these family history categories because having ≥5 affected FDRs is very rare (less than 1 per 100,000 people).

[d] The presented age-specific RR values were calculated by dividing the age-specific RRs for people with ≥1 affected FDR from Table 3 in Fuchs et al.[28] (RRs of 4.63, 3.47, 2.53, 1.69, 1.35, 1.15, and 0.91 for age groups 30-44, 45-49, 50-54, 55-59, 60-64, 65-69, and ≥70 years, respectively) by the overall RR for people with ≥1 affected FDR presented in the same study (RR of 1.72).

Data and assumptions for screening and surveillance

We let colonoscopy screening schedules differ in their start age (30, 35, 40, 45, and 50), interval (1, 2, 3, 5, 7, and 10 years) and end age (75, 80, 85, and 90).

Test characteristics and complication rates of colonoscopy are given in **Table 2**. The sensitivity increases from 75% for small adenomas (≤5 mm) to 85% for medium-sized adenomas (6-9 mm) and to 95% for large adenomas (≥10 mm) and colorectal cancer.[29] The specificity is assumed to be 86%.[30] The lack of specificity reflects the detection of non-adenomatous polyps, which involves unnecessary polypectomy or biopsy. Complications requiring a hospital admission or emergency department visit are assumed to increase exponentially with age.[31-33]

Table 2. Model inputs: Test characteristics, utility loss and costs of colonoscopy screening and treatment.

COLONOSCOPY TEST CHARACTERISTICS				
Sensitivity				
Small adenomas (≤5 mm)	75%[a]			
Medium-sized adenomas (6-9 mm)	85%[a]			
Large adenomas (≥10 mm)	95%[a]			
Colorectal cancer	95%[a]			
Specificity	86%[b]			
Reach	95% reaches the cecum; the reach of the remaining 5% is distributed uniformly over colon and rectum			
Complication rate for positive test				
Serious gastrointestinal event[c]	Age-specific[d]			
Other gastrointestinal event[e]	Age-specific[f]			
Cardiovascular event[g]	Age-specific[h]			
Mortality rate				
Positive test	0.0191 per 1,000[i]			
Negative test	0			
	UTILITY LOSS (QALYs)[j]			
Per colonoscopy	0.0020			
Per complication of colonoscopy				
Serious gastrointestinal event[c]	0.0055			
Other gastrointestinal event[e]	0.0027			
Cardiovascular event[g]	0.0048			
Per LY with CRC care[k,l]	Initial care	Continuing care	Terminal care *Death CRC*	Terminal care *Death other cause*
Stage I CRC	0.12	0.05	0.70	0.05
Stage II CRC	0.18	0.05	0.70	0.05
Stage III CRC	0.24	0.24	0.70	0.24
Stage IV CRC	0.70	0.70	0.70	0.70

Table 2 (continued). Model inputs: Test characteristics, utility loss and costs of colonoscopy screening and treatment.

	COSTS (2014 US$)[m]			
Per colonoscopy				
without polypectomy/biopsy	1,422			
with polypectomy/biopsy	1,699			
Per complication of colonoscopy				
Serious gastrointestinal event[c]	11,142			
Other gastrointestinal event[e]	7,587			
Cardiovascular event[g]	8,453			
Per LY with CRC care[k]	Initial care	Continuing care	Terminal care *Death CRC*	Terminal care *Death other cause*
Stage I CRC	36,883	3,106	64,110	19,331
Stage II CRC	49,475	2,918	63,856	17,429
Stage III CRC	60,033	4,068	67,353	21,620
Stage IV CRC	78,124	12,274	88,749	50,122

QALY = quality-adjusted life year; LY = life year; CRC = colorectal cancer.

[a] The sensitivity of colonoscopy for the detection of adenomas and CRC within the reach of the endoscope was obtained from a systematic review on miss rates observed in tandem colonoscopy studies.[29]

[b] The lack of specificity reflects the detection of non-adenomatous polyps, which leads to unnecessary polypectomy or biopsy.

[c] Serious gastrointestinal events are perforations, gastrointestinal bleeding, or transfusions.

[d] Formula: $1/[\exp(9.27953 - 0.06105 \times \text{Age}) + 1] - 1/[\exp(10.78719 - 0.06105 \times \text{Age}) + 1]$

[e] Other gastrointestinal events are paralytic ileus, nausea and vomiting, dehydration, or abdominal pain.

[f] Formula: $1/[\exp(8.81404 - 0.05903 \times \text{Age}) + 1] - 1/[\exp(9.61197 - 0.05903 \times \text{Age}) + 1]$

[g] Cardiovascular events are myocardial infarction or angina, arrhythmias, congestive heart failure, cardiac or respiratory arrest, syncope, hypotension, or shock.

[h] Formula: $1/[\exp(9.09053 - 0.07056 \times \text{Age}) + 1] - 1/[\exp(9.38297 - 0.07056 \times \text{Age}) + 1]$

[i] Risk of dying from a colonoscopy at age 65 (Warren et al.[31], Gatto et al.[33] and Van Hees et al.[44]).

[j] The loss of quality of life associated with a particular event.

[k] Care for CRC was divided in three clinically relevant phases: the initial, continuing, and terminal care phase. The initial care phase was defined as the first 12 months after diagnosis; the terminal care phase was defined as the final 12 months of life; the continuing care phase was defined as all months in between. In the terminal care phase, we distinguished between CRC patients dying from CRC and CRC patients dying from another cause. For patients surviving less than 24 months, the final 12 months were allocated to the terminal care phase and the remaining months were allocated to the initial care phase.

[l] Utility losses for LYs with initial care were derived from a study by Ness and colleagues.[35] For LYs with continuing care for stage I and II CRC, we assumed a utility loss of 0.05 QALYs; for LYs with continuing care for stage III and IV CRC, we assumed the corresponding utility losses for LYs with initial care. For LYs with terminal care for CRC, we assumed the utility loss for LYs with initial care for stage IV CRC. For LYs with terminal care for another cause, we assumed the corresponding utility losses for LYs with continuing care.

[m] Costs include copayments and patient time costs (i.e. the opportunity costs of spending time on screening or being treated for a complication or CRC), but do not include travel costs, costs of lost productivity, and unrelated health care and non-health care costs in added years of life. We assumed that the value of patient time was equal to the median wage rate in 2014: $17.09 per hour.[45] We assumed that colonoscopies used up 36 hours, serious gastrointestinal complications 192 hours, other gastrointestinal complications 96 hours and cardiovascular complications 120 hours of patient time. Patient time costs associated with CRC care were provided by Yabroff (personal communication), and were calculated using the methodology described in a study by Yabroff and colleagues.[46]

Individuals with adenomas detected and removed at screening were assumed to undergo colonoscopy surveillance according to the current guidelines, and did not return to screening.[34] If the recommended surveillance interval is longer than either the current or any of the past screening intervals, then the surveillance interval is set equal to the minimum of those screening intervals. This ensures that individuals with adenomas detected (i.e. those in surveillance) have colonoscopies at a rate that is as at least as frequently as for those without adenomas detected. We assumed that surveillance continued until 5 years after the end age of screening. Because we wanted to obtain optimal recommendations for individuals following the guideline, adherence to screening and surveillance colonoscopies was assumed to be 100%.

Data and assumptions for costs and utilities

The assumed loss in quality of life due to CRC screening was equivalent to 1.5 day at 0.5 utility per colonoscopy (0.002 QALYs) and 2-4 days at 0.5 utility per complication (0.0027-0.0055 QALYs) (**Table 2**). We also assumed that life years (LYs) with CRC care have a lower quality than those without CRC care.[35]

The cost-effectiveness analyses were conducted from a modified societal perspective. We included both direct medical costs as well as patient time costs. However, direct non-health costs and costs of informal care givers were not included.[36] The costs of colonoscopies were based on 2014 Medicare payment rates and copayments (**Table 2**). For each type of complication, the average payment by Centers for Medicare and Medicaid Services (CMS) was calculated using frequency data on hospitalizations for colonoscopy complications from Elizabeth Drye, MD, SM, and Craig Parzynski, MS, of Yale University (personal communication). Net costs of CRC care were obtained from an analysis of SEER-Medicare linked data[37] (personal communication, Robin Yabroff, PhD, and Martin Brown, PhD, both formerly of the National Cancer Institute). Patient time costs and copayments were added to all of these estimates, which were then updated to 2014 US dollars using the Consumer Price Index.[38]

Outcomes

For each cohort we quantified the effectiveness (i.e., the number of CRC cases prevented, CRC deaths prevented, LYs gained, and QALYs gained) and resources (i.e. number of colonoscopies and costs) of all screening strategies considered, applying the conventional 3% annual discount rate for both.

Base case analysis

For each of the 35 cohorts, we simulated the screening strategies described earlier and determined their costs and effects as compared to no (future) screening. We first compared the costs and effects of 10-yearly and 5-yearly colonoscopy screening for a 50-year old cohort without prior screening for different levels of family history.

Next, we determined the optimal screening strategy for each age group and level of family history. Screening strategies were ranked by increasing costs. Strategies that were

more costly and less effective than another strategy (i.e. strongly dominated strategies) were eliminated from consideration because they were inefficient screening options. Of the remaining strategies, those that were less effective and less costly than another but provided an additional QALY at a higher incremental cost (i.e. weakly dominated strategies) were also eliminated from consideration. For all non-dominated strategies, we calculated the incremental cost-effectiveness ratio (ICER), defined as the additional cost of a specific strategy, divided by its additional clinical benefit (in this case, QALYs gained), compared with the next less expensive strategy (i.e., the strategy with costs closest to, but lower than, the strategy of interest). The optimal screening strategy was then defined as the strategy with an ICER closest to the willingness-to-pay threshold of $100,000 per QALY gained.

The optimal screening strategy was determined in a sequential fashion, starting with the youngest age group for all levels of family history, to allow for appropriate screening history in later age groups (see **Figure 1**). First, we optimized screening for individuals in the youngest age group (30-44 years), using the RR for this age group (**Table 1**). We considered different start ages and let screening continue until age 75, which is currently recommended for the general population. Second, we simulated screening at those ages within 30-44 years that were found to be cost-effective in step 1, and optimized screening for individuals aged 45-49 years, using the RR for this age group. In this case, the start age was only varied (45 or 50 years) if screening was not considered cost-effective at ages 30-44 years. Different intervals were considered for screening from age 45 (or 50) until age 75. Third, we simulated screening at those ages within 30-44 and 45-49 years that were found to be cost-effective in step 1 and 2, and optimized the screening interval at ages 50-75, assuming the RR for age group 50-54 years. The same methodology was applied for age groups 55-59, 60-64, and 65-69. For those aged 70+, different end ages of screening were considered.

Sensitivity analyses

In the sensitivity analyses, we adjusted our base case assumptions in the following way.

Screening history. In the base case, we assumed that previous screening was in accordance with the current number of affected FDRs. In a sensitivity analysis, we considered previous screening to be in line with the guidelines for the general population (i.e. colonoscopy screening at ages 50, 60, and 70).

Relative risk. For some instances, the method of estimating the RRs resulted in very high numbers. As these may not be realistic, we truncated RRs at 10 in a sensitivity analysis.

Cost-effectiveness threshold. Because sometimes a lower willingness-to-pay threshold than $100,000 per QALY may be preferred, we explored the effect of assuming $50,000 per QALY in a sensitivity analysis.

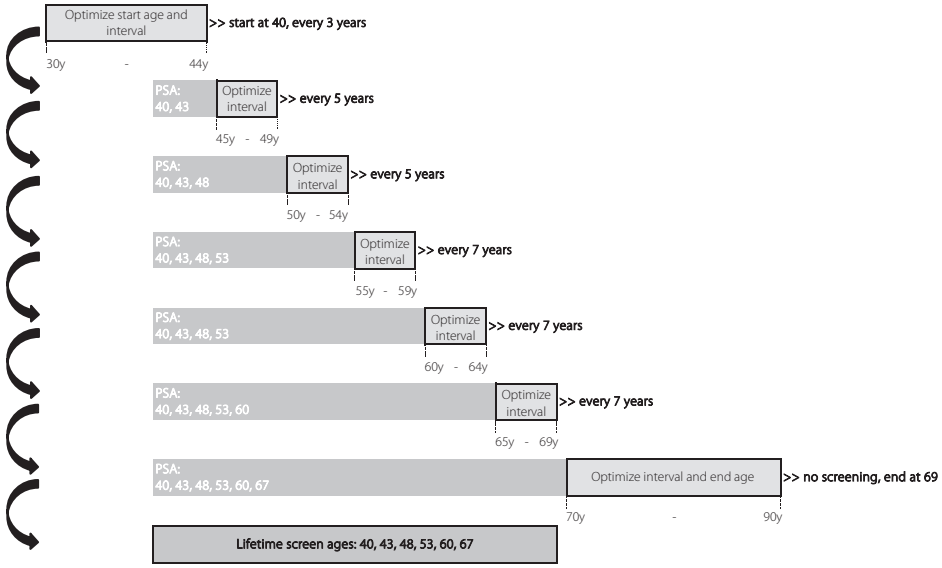


Figure 1. Sequential optimization method, for an example of individuals with one affected FDR (corresponds with the first row in **Table 4**). First, the optimal start age and interval for the youngest cohort (30-year-olds) is determined. The resulting screening ages between ages 30 and 44 are assumed as prior screening for the 45-year-olds, for whom the screening interval from age 45 is optimized. The screening ages until age 49 are then incorporated in the prior screening for 50-year-olds, and so on. For 70-year-olds, the optimal end age of screening is determined. In the figure, the derivation of an optimal screening strategy is given in these subsequent steps (indicated by the black arrows).
PSA = past screening ages (i.e. the assumed screening history for that age group).

RESULTS

For a cohort of 50-year-old previously unscreened individuals with no affected FDRs, no future screening results in 59 CRC cases and 24 CRC deaths per 1,000 simulated individuals. Colonoscopy at ages 50, 60, and 70 prevents 36 of these cases and 19 of these deaths at an expense of almost 3,400 colonoscopies (**Table 3**). Screening every 5 years instead of every 10 years requires 1,711 additional colonoscopies to prevent 4.4 additional CRC cases and 1.5 additional CRC deaths. This implies that over 1,000 additional colonoscopies are needed to prevent one additional CRC death. For individuals with 1, 2, 3, or ≥4 FDRs, this 'number needed to screen' was 304, 192, 135, and 75, respectively. In terms of cost-effectiveness, replacing 10-yearly screening by 5-yearly screening would cost $186,000 per QALY gained in people without any affected FDRs, and $26,000, $9,000, and $1,000 per QALY gained in people with 1, 2, and 3 affected FDRs, respectively. It is potentially cost saving in people with 4 or more FDRs.

Table 3. Costs and effects of screening a previously unscreened 50-year-old cohort with colonoscopy until age 70 for different levels of familial risk at age 50, using either a 10- or a 5-year screening interval, as compared to no screening (effectiveness and costs per 1,000 simulated persons).

No. of affected FDRs at age 50	Screening interval	EFFECTS COMPARED TO NO SCREENING				COSTS COMPARED TO NO SCREENING[c] (*$1,000)				INCREMENTAL EFFECTS AND COSTS WHEN 10-YEAR INTERVAL IS REPLACED BY 5-YEAR INTERVAL[c]		
		Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b,c]	Screening and surveillance	Complications	CRC diagnosis and care[d]	Total costs	QALYs gained	Total costs (*$1,000)	Costs per QALY gained (*$1,000)
0	10	3,381	36	19	104	3,860	60	-1,488	2,433	+9 (+8%)	+1,621 (+67%)	185
	5	5,092	40	20	113	5,689	77	-1,712	4,054			
1	10	4,613	113	60	343	5,297	144	-4,778	663	+26 (+7%)	+659 (+99%)	26
	5	5,701	123	64	369	6,452	162	-5,292	1,322			
2	10	5,201	171	92	544	6,086	196	-7,393	-1,111	+25 (+5%)[e]	+216 (+24%)	9
	5	5,803	179	95	568	6,759	209	-7,863	-895			
3	10	5,589	229	128	785	6,721	242	-10,340	-3,378	+18 (+2%)[e]	+11 (+0%)	1
	5	5,857	235	130	803	7,042	249	-10,657	-3,367			
≥4	10	6,104	348	226	1,557	7,768	326	-17,955	-9,860	+2 (+0%)[e]	-12 (-0%)	Potentially cost-saving
	5	6,117	348	226	1,559	7,785	327	-17,985	-9,873			

FDR = first-degree relative; CRC = colorectal cancer; QALY = quality-adjusted life year.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] The impact of screening on quantity and quality of life incorporated in one measure, i.e. the net health benefit of screening.

[c] Costs and QALYs were discounted with 3%.

[d] Screening prevents costs by preventing LYs with CRC care and induces costs by adding LYs with CRC care. The net effect can be a reduction in costs (negative values) or an increase in costs (positive values).

[e] For people with 2 or more affected FDRs, the number of QALYs gained by replacing 10-yearly by 5-yearly colonoscopy is lower than for people with 1 affected FDR. This is because the adenoma risk for people with 2 or more FDRs is very high, and a large share of the simulated cohort will therefore go into surveillance. The screening interval is only applied to people without a history of adenomas, and not for those in surveillance.

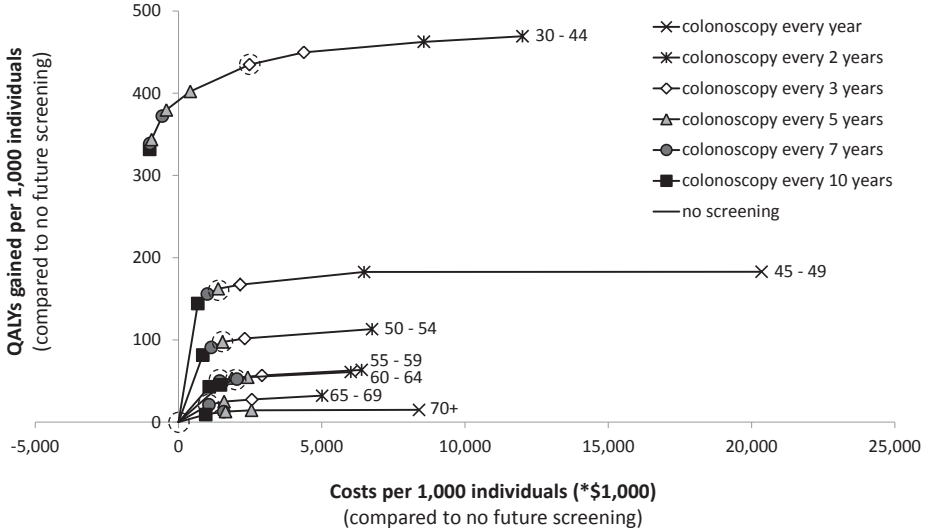


Figure 2. Cost-efficiency frontiers for people with one affected FDR, by age. The costs, health effects and ICER of every strategy in the figure can be found in **Appendix Tables 1a - 1g**. If the same symbols appear in a single frontier, strategies begin or end at different ages (see the Appendix tables for the exact strategies). Strategies that are cost-effective under a willingness-to-pay threshold of $100,000 per QALY gained are marked.

For example, the "50 - 54" frontier includes strategies with different intervals for screening at ages 50 to 75 in individuals with a relative risk of 3.00 (**Table 1**) and with past screening at ages 40, 43, and 48 years (**Figure 1**).

Cost-effective screening for individuals with one affected FDR

For individuals with a single affected FDR, optimization of screening for the youngest age group showed that 3-yearly screening ideally starts at age 40 (**Table 4**). After two consecutive negative colonoscopies at ages 40 and 43, it is cost-effective to lengthen the screening interval to 5 years. From ages 55 to 70, the screening interval can be 7 years for those with a screening history of only negative colonoscopies. For illustrative purposes, **Figure 2** shows the age-specific cost-efficiency frontiers in which the optimal strategies are marked.

Cost-effective screening for individuals with two or more affected FDRs

Individuals who already have 2 affected FDRs at young age should start 3-yearly screening at age 35. After subsequent negative colonoscopies, this interval is preferably extended to 5 years at age 55, and to 7 years at age 70. For individuals with 3 or ≥4 affected FDRs, 2-yearly screening is recommended from age 35 or 30, respectively. For individuals with 3 affected FDRs, the interval can be lengthened to 3, 5, and 7 years at age 45, 60, and 70 respectively. For those with ≥4 affected FDRs, intensive screening remains cost-effective at older ages, and the interval can only be extended to 3 years at age 70.

Table 4. Optimal colonoscopy screening intervals under base case assumptions (threshold = $100,000 per QALY gained) by age group and number of affected FDRs.[a]

		AGE GROUP OF PERSON TO BE SCREENED									
		30-34	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**	**60-64**	**65-69**	**70-74**	**75+**
FAMILY HISTORY	**1 affected FDR**	-	-	3	5	5	7	7	7	-[c]	-
	2 affected FDRs	-	3	3	3	3[b]	5	5	5[b]	7	-
	3 affected FDRs	-	2[b]	2[b]	3	3	3	5	5	7[b]	-
	≥4 affected FDRs	2	2	2	2	2	2	2	2	3	-

FDR = first-degree relative

[a] The effects, costs and ICER of all efficient strategies for individuals with 1, 2, 3, and ≥4 affected FDRs are given in **Appendix Tables 1**, **2**, **3**, and **4**, respectively.

[b] The ICER is just below the willingness-to-pay threshold of $100,000 per QALY gained (i.e., between $90,000 and $100,000).

[c] The next ICER is just above the willingness-to-pay threshold of $100,000 per QALY gained (i.e., between $100,000 and $110,000).

Table 5. Optimal colonoscopy screening intervals in sensitivity analyses (threshold = $100,000 per QALY gained, unless stated otherwise) by age group and number of affected FDRs.[a]

		AGE GROUP OF PERSON TO BE SCREENED									
		30-34	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**	**60-64**	**65-69**	**70-74**	**75+**
FAMILY HISTORY	RRs truncated at 10[b]										
	3 affected FDRs	-	3	3	3[e]	2	3	5[e]	5	7[d]	-
	≥4 affected FDRs	-	3	3	2[d]	3	2[d]	2	2	3[e]	-
	Threshold of $50,000 per QALY gained										
	1 affected FDR	-	-	5	7[e]	7	7[d]	10	10[e]	-	-
	2 affected FDRs	-	-	3	5	5	7[e]	7	10[e]	7[d]	-
	3 affected FDRs	-	3[d]	3[d]	5[e]	3[d]	5	5	5[d]	-[e]	-
	≥4 affected FDRs	-	2[e]	2[e]	3	2	3	3	3[d]	-	-
	Prior screening as in general population[c]										
	1 affected FDR	-	-	3	3	3[d]	5	7[e]	7	3	-
	2 affected FDRs	-	3	3	3	3	3	3	3	3	-
	3 affected FDRs	-	2[d]	2[d]	2	3[e]	3	3	2[d]	3	-
	≥4 affected FDRs	2	2	2	2	2	2	2	2	2	-

FDR = first-degree relative; RR = relative risk (as compared to the general population); QALY = quality-adjusted life year.

[a] The effects, costs and ICER of all efficient strategies for individuals with 1, 2, 3, and ≥4 affected FDRs are given in **Appendix Tables 1**, **2**, **3**, and **4**, respectively.

[b] For none of the age groups, having less than 3 FDRs is associated with a RR above 10.

[c] Prior screening was assumed to be colonoscopy screening at ages 50, 60, and 70 instead of what is optimal for that family history category.

[d] The ICER is just below the willingness-to-pay threshold (i.e., between $90,000 and $100,000 when the willingness to pay was $100,000 per QALY, and between $40,000 and $50,000 when the willingness to pay was $50,000 per QALY).

[e] The next ICER is just above the willingness-to-pay threshold (i.e., between $100,000 and $110,000 when the willingness to pay was $100,000 per QALY, and between $50,000 and $60,000 when the willingness to pay was $50,000 per QALY).

Sensitivity analyses

Results of the sensitivity analyses are shown in **Table 5**. When relative risks were truncated at 10, individuals with ≥3 affected FDRs preferably start off with 3-yearly instead of 2-yearly colonoscopy screening. However, it was cost-effective to switch to a 2-year interval at a later age. From age 55, optimal screening strategies were identical to the base case analysis.

When we assumed that people had prior screening according to what is recommended for the general population instead of to their family history level, screening intervals did not tend to lengthen with age.

As expected, a lower cost-effectiveness threshold resulted in less intensive screening.

DISCUSSION

This study confirms that it is effective and cost-effective to screen people with affected FDRs more often than people in the general population. The more affected FDRs an individual has, the earlier the start age of screening and the shorter the preferred screening interval. For individuals with a single affected FDR, who make up approximately 92% of people with at least one affected FDR,[5] the optimal screening interval gradually increased from 3 years at age 40 to 7 years at age 55. This increase suggests that the benefits of age-specific CRC screening guidelines for people with a positive family history, but with persistent negative screening results, may be substantial. As the optimal screening interval did not (significantly) lengthen with age in individuals who had prior screening as recommended for the general population, it is crucial that intensified screening is offered as soon as an FDR is diagnosed with CRC. An increase of the screening interval can then be considered at a later stage, when an individual has had several negative colonoscopies.

We modeled an increased risk of developing CRC as an increased risk of developing adenomas. Although the duration distribution of a single adenoma to progress to CRC was assumed to be independent of family history, we found that it is cost-effective to screen individuals at increased risk at shorter intervals than the general population. There are two reasons for this finding. The first one is that at every screening, a random percentage of adenomas is missed due to a lack of sensitivity, and this percentage translates to a larger absolute number of missed adenomas in higher-risk populations. Most of these missed adenomas will be picked up by subsequent screenings, but some may progress to cancer before being detected. The second reason is that, although in every risk group, the same (small) percentage of adenomas is fast-growing, the absolute number of fast-growing adenomas is higher in those at increased risk. Both mechanisms underscore the need for screening with shorter intervals in those at higher risk, which has been explained more extensively elsewhere.[39]

Offering individuals with one affected FDR colonoscopy screening at a 3-year interval may seem aggressive, and is more intensive than existing guidelines.[40] However, individuals who already have an affected FDR at young age are more likely to have multiple affected

FDRs at a later age. This is reflected in the familial RR for CRC being much higher for younger individuals than it is for older individuals.[28] Intensified screening at young age, with a subsequent lengthening of the screening interval after consecutive negative findings, could therefore be considered a very reasonable option for those with affected FDRs.

As the optimal screening interval varied both with age and number of affected FDRs, guidelines could be improved by a more detailed grouping of family history. An earlier study already showed that intervals of less than 5 years may be appropriate for those with multiple affected FDRs.[9] The fact that this has not been incorporated in guidelines yet, may be because it is considered too complex by policy makers. However, it only involves a small group of individuals and does not affect the general population. The age-dependency could even be simplified, by e.g. lengthening the interval only at age 55, and only for individuals with 1-3 affected FDRs (e.g. from 5 to 7 years for those with 1 affected FDR, and from 3 to 5 years for those with 2-3 affected FDRs). Moreover, when an individual is diagnosed with CRC, the FDRs will often be identified quite easily. At genetic counseling, the CRC risk of these FDRs and the associated optimal screening strategy could be determined using e.g. an individual risk score such as the Framingham risk score for cardiovascular disease. In the Netherlands, surveillance of individuals with detected adenomas is already based on an individual risk score.[41]

Although we presented results for different levels of family history and age groups in a similar way, it should be noted that having 3 or more affected FDRs is rare (i.e., less than 1% of all people with a family history), especially at young age.[5] It cannot be ruled out that these people have an undiagnosed syndrome that increases their CRC risk significantly. However, as long as such a syndrome has not been diagnosed yet, screening these people more often based on their number of affected FDRs is probably beneficial.

In general, US guidelines recommend that screening begins at age 40 for individuals with one FDR diagnosed with CRC below age 60 or two or more FDRs diagnosed at any age.[40] In contrast with an earlier study, we found that for individuals with ≥2 affected FDRs, the preferred start age of screening is below the age of 40.[9] One US guideline states that for individuals with one FDR diagnosed with CRC at age ≥60, multiple negative colonoscopies may support lengthening the colonoscopy interval.[42] We showed that such lengthening is cost-effective, and that it may also be recommended for individuals with a more pronounced family history of CRC.

In this study, we only considered the number of FDRs diagnosed with CRC to estimate familial CRC risk. Although to a smaller extent, combinations of affected second- and third-degree relatives can also increase colorectal cancer risk significantly.[5] The same is true for FDRs with adenomas detected instead of cancer.[43] The inclusion of such alternatives would, from a computational perspective, only imply considering slightly lower relative risks. The optimal screening intervals may then be longer, but the finding that age-specific screening guidelines are likely to result in large benefits would still apply.

This study also has its limitations. First, using the number of affected FDRs gives an indication of an individual's family history, but is dependent on his or her family size. For an

individual with a small family, one affected FDR is more telling than for an individual with a large family. For individuals without siblings and offspring, it would not be possible to have 3 or more affected FDRs, regardless of the level of familial CRC risk. Second, in the absence of true age-specific estimates of RRs for all levels of affected FDRs, we estimated these RRs ourselves. As in some cases our approach ended up in relatively high values, we also truncated the RRs in a sensitivity analysis. Finally, we assumed that an individual's level of family history remains constant over time, although in real life, it is likely to increase as an individual ages. Although exploring all possible life histories was not feasible, we did consider a scenario with prior screening according to what is recommended for the general population. The results of this analysis show the screening interval with which intensified screening should start off with, in case the family history is revealed at that age. At older ages, the optimal screening interval then lies between the optimal interval found in that analysis and the one found in the base-case analysis.

In summary, we have shown that it is cost-effective to offer individuals with a positive family history of CRC more intensive colonoscopy screening than people in the general population, especially at young age. Furthermore, for individuals with a constant level of family history over time, it is cost-effective to gradually increase the screening interval after several subsequent negative colonoscopies. If no adenomas develop, it is unlikely that these individuals are affected by genetic predisposition, so continuing intensive colonoscopy screening would provide little or no additional health benefit and is clearly not cost-effective.

ACKNOWLEDGMENTS

This work resulted from a collaboration between the Cancer Intervention and Surveillance Modeling Network (CISNET) and the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. The authors would like to thank Elizabeth Drye, MD, SM, and Craig Parzynski, MS, of Yale University, who provided us with frequency data on hospitalizations for colonoscopy complications.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66:7-30.
2. Fearnhead NS, Wilding JL, Bodmer WF. Genetics of colorectal cancer: hereditary aspects and overview of colorectal tumorigenesis. Br Med Bull 2002;64:27-43.
3. Butterworth AS, Higgins JP, Pharoah P. Relative and absolute risk of colorectal cancer for individuals with a family history: a meta-analysis. Eur J Cancer 2006;42:216-27.
4. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. N Engl J Med 2003;348:919-32.
5. Taylor DP, Burt RW, Williams MS, et al. Population-based family history-specific risks for colorectal cancer: a constellation approach. Gastroenterology 2010;138:877-85.
6. U.S. Preventive Services Task Force. Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2008;149:627-37.
7. Levin B, Lieberman DA, McFarland B, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. Gastroenterology 2008;134:1570-95.
8. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, et al. Evaluating test strategies for colorectal cancer screening: a decision analysis for the U.S. Preventive Services Task Force. Ann Intern Med 2008;149:659-69.
9. Wilschut JA, Steyerberg EW, van Leerdam ME, et al. How much colonoscopy screening should be recommended to individuals with various degrees of family history of colorectal cancer? Cancer 2011;117:4166-74.
10. Henrikson NB, Webber EM, Goddard KA, et al. Family history and the natural history of colorectal cancer: systematic review. Genet Med 2015.
11. Rutter CM, Johnson EA, Feuer EJ, et al. Secular trends in colon and rectal cancer relative survival. J Natl Cancer Inst 2013;105:1806-13.
12. SEER*Stat Database: Incidence–SEER 9 Regs Public-Use NS-. Surveillance, Epidemiology, and End Results (SEER) Program. November 2003 ed.
13. Blatt LJ. Polyps of the colon and rectum: Incidence and distribution. Dis Colon Rectum 1961;4:277-282.
14. Arminski TC, McLean DW. Incidence and Distribution of Adenomatous Polyps of the Colon and Rectum Based on 1,000 Autopsy Examinations. Dis Colon Rectum 1964;7:249-61.
15. Bombi JA. Polyps of the colon in Barcelona, Spain. An autopsy study. Cancer 1988;61:1472-6.
16. Chapman I. Adenomatous polypi of large intestine: incidence and distribution. Ann Surg 1963;157:223-6.
17. Clark JC, Collan Y, Eide TJ, et al. Prevalence of polyps in an autopsy series from areas with varying incidence of large-bowel cancer. Int J Cancer 1985;36:179-86.
18. Jass JR, Young PJ, Robinson EM. Predictors of presence, multiplicity, size and dysplasia of colorectal adenomas. A necropsy study in New Zealand. Gut 1992;33:1508-14.
19. Johannsen LG, Momsen O, Jacobsen NO. Polyps of the large intestine in Aarhus, Denmark. An autopsy study. Scand J Gastroenterol 1989;24:799-806.
20. Rickert RR, Auerbach O, Garfinkel L, et al. Adenomatous lesions of the large bowel: an autopsy survey. Cancer 1979;43:1847-57.
21. Vatn MH, Stalsberg H. The prevalence of polyps of the large intestine in Oslo: an autopsy study. Cancer 1982;49:819-25.
22. Williams AR, Balasooriya BA, Day DW. Polyps and cancer of the large bowel: a necropsy study in Liverpool. Gut 1982;23:835-42.

23. Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. Lancet 2010;375:1624-33.
24. Hardcastle JD, Chamberlain JO, Robinson MH, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet 1996;348:1472-7.
25. Jorgensen OD, Kronborg O, Fenger C. A randomised study of screening for colorectal cancer using faecal occult blood testing: results after 13 years and seven biennial screening rounds. Gut 2002;50:29-32.
26. Lansdorp-Vogelaar I, van Ballegooijen M, Boer R, et al. A novel hypothesis on the sensitivity of the fecal occult blood test: Results of a joint analysis of 3 randomized controlled trials. Cancer 2009;115:2410-9.
27. Mandel JS, Church TR, Ederer F, et al. Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. J Natl Cancer Inst 1999;91:434-7.
28. Fuchs CS, Giovannucci EL, Colditz GA, et al. A prospective study of family history and the risk of colorectal cancer. N Engl J Med 1994;331:1669-74.
29. van Rijn JC, Reitsma JB, Stoker J, et al. Polyp miss rate determined by tandem colonoscopy: a systematic review. Am J Gastroenterol 2006;101:343-50.
30. Schroy PC, 3rd, Coe A, Chen CA, et al. Prevalence of advanced colorectal neoplasia in white and black patients undergoing screening colonoscopy in a safety-net hospital. Ann Intern Med 2013;159:13-20.
31. Warren JL, Klabunde CN, Mariotto AB, et al. Adverse events after outpatient colonoscopy in the Medicare population. Ann Intern Med 2009;150:849-57, W152.
32. van Hees F, Zauber AG, Klabunde CN, et al. The Appropriateness of More Intensive Colonoscopy Screening Than Recommended in Medicare Beneficiaries - A Modeling Study. JAMA Intern Med 2014;174:1568-1576.
33. Gatto NM, Frucht H, Sundararajan V, et al. Risk of perforation after colonoscopy and sigmoidoscopy: a population-based study. J Natl Cancer Inst 2003;95:230-6.
34. Lieberman DA, Rex DK, Winawer SJ, et al. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 2012;143:844-57.
35. Ness RM, Holmes AM, Klein R, et al. Utility valuations for outcome states of colorectal cancer. Am J Gastroenterol 1999;94:1650-7.
36. Gold MR, Siegel JE, Russell LB, et al. Cost-effectiveness in Health and Medicine. New York: Oxford University Press, 1996.
37. Yabroff KR, Lamont EB, Mariotto A, et al. Cost of care for elderly cancer patients in the United States. J Natl Cancer Inst 2008;100:630-41.
38. The United States Department of Labor. Bureau of Labor Statistics. Consumer Price Index.
39. O'Mahony JF, van Rosmalen J, Mushkudiani NA, et al. The influence of disease risk on the optimal time interval between screens for the early detection of cancer: a mathematical approach. Med Decis Making 2015;35:183-95.
40. Ramsey SD, Wilschut J, Boer R, et al. A decision-analytic evaluation of the cost-effectiveness of family history-based colorectal cancer screening programs. Am J Gastroenterol 2010;105:1861-9.
41. Nederlandse Vereniging van Maag-, Darm- en Leverartsen (NVMDL). Nederlandse Richtlijn Coloscopie Surveillance, 2013.
42. Colorectal Cancer Screening. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines): National Comprehensive Cancer Network, 2016.
43. Nakama H, Zhang B, Fukazawa K, et al. Family history of colorectal adenomatous polyps as a risk factor for colorectal cancer. Eur J Cancer 2000;36:2111-4.

Appendix Table 1a. Costs and effects of cost-effective screening strategies for people with 1 FDR, aged 30-44 years (relative risk = 5.49), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY		EFFECTS				COSTS[b] (*1,000)	ICER
Begin age	Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
50	10	5,114	184	100	332	-$1,006	Reference
50	7	5,322	188	102	339	-$1,000	$764
50	5	5,558	191	103	344	-$940	$12,586
45	7	6,156	195	105	372	-$565	$13,101
45	5	6,536	198	106	380	-$422	$19,194
40	5	7,520	202	108	402	$412	$37,113
40	3	11,748	230	114	435	$2,477	$63,176
35	3	13,540	234	116	450	$4,375	$127,429
35	2	19,767	249	118	462	$8,562	$329,289
30	2	22,488	253	119	469	$12,005	$502,885

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 1b. Costs and effects of cost-effective screening strategies for people with 1 FDR, aged 45-49 years (relative risk = 4.12), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
10	3,495	70	34	144	$675	$4,679
7	3,941	73	35	156	$1,007	$28,716
5	4,354	75	36	162	$1,386	$58,808
3	5,077	77	36	167	$2,153	$156,342
2	10,133	91	39	183	$6,478	$280,702
1	25,276	105	41	183	$20,348	$42,909,079

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 1c. Costs and effects of cost-effective screening strategies for people with 1 FDR, aged 50-54 years (relative risk = 3), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
10	2,452	40	19	82	$850	$10,408
7	2,851	43	20	91	$1,144	$32,155
5	3,333	45	21	97	$1,543	$59,794
3	4,042	47	22	102	$2,312	$183,136
2	8,754	57	24	113	$6,757	$389,235

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 1d. Costs and effects of cost-effective screening strategies for people with 1 FDR, aged 55-59 years (relative risk = 2), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
10	1,898	21	10	43	$1,072	$24,950
7	2,341	24	11	50	$1,442	$53,334
5	2,762	25	12	54	$1,864	$114,289
3	3,745	28	12	57	$2,913	$321,954
2	7,245	34	14	63	$6,393	$522,666

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 1e. Costs and effects of cost-effective screening strategies for people with 1 FDR, aged 60-64 years (relative risk = 1.6), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
10	1,995	19	9	45	$1,467	$32,497
7	2,540	21	10	52	$2,037	$78,058
5	2,915	23	11	55	$2,418	$173,738
2	6,264	28	12	61	$6,013	$568,574

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 1f. Costs and effects of cost-effective screening strategies for people with 1 FDR, aged 65-69 years (relative risk = 1.36), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
7	1,206	9	4	21	$1,065	$50,292
5	1,701	11	5	25	$1,592	$136,564
3	2,509	12	6	27	$2,560	$395,336
2	4,666	16	7	32	$5,013	$518,476

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 1g. Costs and effects of cost-effective screening strategies for people with 1 FDR, aged ≥70 years (relative risk = 1.08), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY		EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	End age	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening		0	0	0	0	$0	Reference
10	75	551	4	2	9	$949	$102,041
7	75	1,126	5	2	13	$1,575	$182,012
5	75	1,148	5	2	13	$1,642	$425,004
5	80	2,046	7	3	14	$2,549	$574,994
1	75	6,644	11	4	15	$8,393	$12,498,621

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 2a. Costs and effects of cost-effective screening strategies for people with 2 FDRs, aged 30-44 years (relative risk = 8.66), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY		EFFECTS				COSTS[b] (*1,000)	ICER
Begin age	Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
50	7	5,555	253	145	507	-$2,625	Reference
50	5	5,616	254	146	510	-$2,621	$1,422
45	7	6,467	264	150	561	-$2,316	$5,940
45	5	6,607	266	151	566	-$2,278	$7,368
40	5	7,583	271	153	602	-$1,567	$20,056
40	3	11,630	311	163	652	$39	$32,083
35	3	13,399	317	165	676	$1,838	$73,695
35	2	19,643	340	168	700	$5,792	$168,812
30	2	22,340	345	170	713	$9,165	$248,190
30	1	44,179	373	174	718	$25,957	$3,328,068

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 2b. Costs and effects of cost-effective screening strategies for people with 2 FDRs, aged 45-49 years (relative risk = 6.49), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
10	2,210	58	29	118	-$42	Reference
7	2,433	60	30	126	$88	$16,369
5	2,626	62	30	131	$235	$30,951
3	2,926	63	30	134	$522	$84,472
2	8,141	81	34	160	$4,939	$167,258
1	23,297	104	37	172	$18,522	$1,127,816

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 2c. Costs and effects of cost-effective screening strategies for people with 2 FDRs, aged 50-54 years (relative risk = 4.73), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
10	2,105	44	22	93	$361	$3,881
7	2,384	47	23	102	$521	$17,814
5	2,602	49	23	107	$703	$38,385
3	2,997	51	24	111	$1,096	$98,358
2	6,815	64	26	132	$4,127	$142,155
1	19,754	81	29	140	$16,481	$1,663,780

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 2d. Costs and effects of cost-effective screening strategies for people with 2 FDRs, aged 55-59 years (relative risk = 3.16), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
10	1,724	25	12	49	$747	$15,234
7	1,976	27	13	56	$912	$24,772
5	2,204	29	13	59	$1,111	$54,755
3	2,680	31	14	63	$1,584	$135,877
2	6,158	39	15	76	$4,945	$256,912
1	16,847	50	17	78	$15,840	$5,465,006

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 2e. Costs and effects of cost-effective screening strategies for people with 2 FDRs, aged 60-64 years (relative risk = 2.52), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
10	1,407	15	7	31	$866	$27,891
7	1,670	18	8	37	$1,062	$34,885
5	1,897	19	9	40	$1,273	$68,198
3	2,312	20	9	42	$1,732	$191,228
2	5,329	26	10	52	$4,922	$328,340
1	13,843	34	12	53	$14,210	$13,203,619

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 2f. Costs and effects of cost-effective screening strategies for people with 2 FDRs, aged 65-69 years (relative risk = 2.15), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
7	1,298	11	5	25	$1,032	$41,275
5	1,567	13	6	28	$1,295	$91,212
2	3,892	19	7	38	$3,741	$247,718
1	10,184	24	8	38	$11,027	$11,392,666

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 2g. Costs and effects of cost-effective screening strategies for people with 2 FDRs, aged 70+ years (relative risk = 1.7), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY		EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	End age	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening		0	0	0	0	$0	Reference
7	75	1,251	8	4	20	$1,719	$88,154
2	75	2,901	12	4	26	$3,559	$299,006
2	80	4,532	14	6	27	$5,186	$1,270,327

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 3a. Costs and effects of cost-effective screening strategies for people with 3 FDRs, aged 30-44 years (relative risk = 12.74), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY		EFFECTS				COSTS[b] (*1,000)	ICER
Begin age	Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
50	5	5,724	306	188	695	-$4,242	Reference
45	7	6,730	320	195	774	-$4,106	$1,706
45	5	6,777	321	195	777	-$4,099	$2,778
45	3	9,659	365	206	831	-$3,542	$10,242
40	3	11,451	378	211	895	-$2,417	$17,749
35	3	13,203	385	213	931	-$725	$46,433
35	2	19,488	418	219	968	$2,951	$98,111
30	2	22,153	424	221	990	$6,245	$154,019
30	1	43,927	464	227	1,008	$22,699	$904,046

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 3b. Costs and effects of cost-effective screening strategies for people with 3 FDRs, aged 45-49 years (relative risk = 9.55), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
10	2,227	56	27	109	$0	Reference
7	2,418	58	27	116	$114	$16,461
5	2,547	59	27	119	$224	$36,134
3	2,708	59	28	122	$399	$80,612
2	2,821	59	28	122	$551	$291,643
1	17,721	93	33	150	$13,291	$449,190

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 3c. Costs and effects of cost-effective screening strategies for people with 3 FDRs, aged 50-54 years (relative risk = 6.96), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
10	1,724	34	15	55	$369	$6,729
7	1,850	35	16	59	$428	$13,377
5	1,939	36	16	62	$488	$24,528
3	2,054	37	16	63	$597	$64,655
2	2,140	37	16	64	$701	$222,261
1	15,325	61	20	83	$13,020	$625,615

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 3d. Costs and effects of cost-effective screening strategies for people with 3 FDRs, aged 55-59 years (relative risk = 4.65), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
10	1,986	29	13	52	$810	$15,490
7	2,146	31	14	58	$909	$17,684
5	2,293	32	14	61	$1,031	$40,143
3	2,526	33	14	64	$1,265	$86,468
2	2,664	33	14	64	$1,417	$294,736
1	13,347	50	17	75	$11,936	$928,522

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 3e. Costs and effects of cost-effective screening strategies for people with 3 FDRs, aged 60-64 years (relative risk = 3.71), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
7	1,794	20	9	34	$1,077	$31,467
5	1,918	21	9	36	$1,179	$45,183
3	2,112	22	9	38	$1,382	$121,584
2	2,336	22	9	39	$1,630	$289,168
1	11,136	34	11	47	$11,002	$1,201,491

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 3f. Costs and effects of cost-effective screening strategies for people with 3 FDRs, aged 65-69 years (relative risk = 3.16), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
7	1,624	15	6	26	$1,238	$47,490
5	1,763	16	7	28	$1,360	$53,457
3	1,970	17	7	30	$1,578	$139,720
2	2,140	17	7	30	$1,793	$693,721
1	8,613	26	9	37	$8,985	$1,055,080

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 3g. Costs and effects of cost-effective screening strategies for people with 3 FDRs, aged ≥70 years (relative risk = 2.5), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY		EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	End age	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening		0	0	0	0	$0	Reference
7	75	1,656	12	5	22	$2,134	$96,786
3	75	1,860	12	5	23	$2,366	$186,368
3	80	3,293	16	6	26	$3,685	$531,436
2	80	3,529	16	6	26	$3,984	$1,667,134
1	75	6,583	18	6	28	$7,866	$1,669,109

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 4a. Costs and effects of cost-effective screening strategies for people with ≥4 FDRs, aged 30-44 years (relative risk = 28.99), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY		EFFECTS				COSTS[b] (*1,000)	ICER
Begin age	Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
45	3	8,726	425	305	1,449	-$8,187	Reference
40	3	10,578	447	314	1,589	-$7,768	$2,995
35	3	12,350	458	319	1,671	-$6,434	$16,156
35	2	18,711	526	332	1,758	-$3,721	$31,312
30	2	21,344	534	335	1,811	-$659	$58,409
30	1	42,979	615	347	1,876	$14,735	$234,257

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 4b. Costs and effects of cost-effective screening strategies for people with ≥4 FDRs, aged 45-49 years (relative risk = 21.72), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
10	845	31	13	34	-$100	Reference
7	853	31	13	34	-$99	$2,241
5	859	31	14	35	-$96	$15,175
3	864	31	14	35	-$91	$28,958
2	868	31	14	35	-$88	$54,504
1	16,020	92	22	114	$12,053	$154,166

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 4c. Costs and effects of cost-effective screening strategies for people with 4 FDRs, aged 50-54 years (relative risk = 15.84), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
7	1,179	34	15	41	$54	$1,321
5	1,183	35	15	41	$55	$7,971
3	1,187	35	15	41	$59	$39,060
2	1,191	35	15	41	$63	$59,534
1	13,649	81	21	105	$10,239	$160,074

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 4d. Costs and effects of cost-effective screening strategies for people with 4 FDRs, aged 55-59 years (relative risk = 10.58), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
7	1,531	34	14	43	$385	$8,983
5	1,535	34	14	43	$388	$10,747
3	1,540	34	14	43	$391	$24,483
2	1,544	34	14	43	$395	$58,840
1	12,378	63	18	85	$10,441	$242,229

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 4e. Costs and effects of cost-effective screening strategies for people with 4 FDRs, aged 60-64 years (relative risk = 8.45), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
5	1,720	32	13	48	$684	$14,349
3	1,726	32	13	48	$689	$42,699
2	1,733	32	13	48	$695	$63,187
1	9,981	54	16	82	$8,438	$228,621

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 4f. Costs and effects of cost-effective screening strategies for people with 4 FDRs, aged 65-69 years (relative risk = 7.2), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY	EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening	0	0	0	0	$0	Reference
5	1,877	30	12	53	$1,016	$19,199
3	1,881	30	12	53	$1,019	$49,298
2	1,885	30	12	53	$1,024	$73,751
1	8,570	45	15	78	$7,943	$273,061

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Appendix Table 4g. Costs and effects of cost-effective screening strategies for people with 4 FDRs, aged ≥70 years (relative risk = 5.7), as compared to no screening. Results per 1,000 simulated persons. The optimal strategy (under a threshold of $100,000 per QALY gained) is surrounded by dashed lines.

STRATEGY		EFFECTS				COSTS[b] (*1,000)	ICER
Screening interval	End age	Colonoscopies[a]	CRC cases prevented	CRC deaths prevented	QALYs gained[b]		
No screening		0	0	0	0	$0	Reference
7	75	2,122	28	11	55	$2,537	$46,161
3	75	2,133	28	11	55	$2,546	$56,845
2	75	2,141	28	11	55	$2,553	$159,292
1	75	6,320	39	13	74	$6,482	$207,467
1	80	8,961	43	15	75	$8,964	$2,582,406

CRC = colorectal cancer; QALY = quality-adjusted life year; ICER = incremental cost-effectiveness ratio.

[a] Number of colonoscopies performed for screening, surveillance and diagnosis of CRC, minus the number of clinical CRC diagnoses in case of no screening.

[b] QALYs and costs were discounted with 3%.

Chapter 10

Exploration of the benefit of risk-stratified colorectal cancer screening based on common genetic variants – current status and future potential

Steffie K. Naber, Suman Kundu, Karen M. Kuntz, W. David Dotson, Marc S. Williams, Ned Calonge, Doris T. Zallen, Theodore G. Ganiats, Evelyn P. Whitlock, Elizabeth M. Webber, Katrina A.B. Goddard, Nora B. Henrikson, Marjolein van Ballegooijen, A. Cecile J.W. Janssens, Ann G. Zauber, Iris Lansdorp-Vogelaar

To be submitted.

ABSTRACT

Importance: Although uniform colonoscopy screening reduces colorectal cancer (CRC) mortality, screening based on polygenic risk may be more efficient.

Objective: To investigate whether risk-stratified CRC screening based on polygenic risk is a cost-effective alternative to current uniform screening, and if not, under what conditions it would be.

Design, Setting and Participants: The MISCAN-Colon model was used to simulate a hypothetical cohort of US 40-year-olds willing to participate in a colonoscopy-based screening program, regardless of whether it involves uniform or risk-stratified screening.

Exposures: Uniform screening was modeled as colonoscopy screening at ages 50, 60, and 70. Prior to risk-stratified screening, individuals were offered a polygenic test with an area under the ROC curve (AUC) of its current projected value of 0.60, and of potential future values 0.65, 0.70, 0.75, and 0.80. Based on the results of the polygenic test, the population was subdivided into 60 risk groups with relative risk (RR) varying from <0.1 to >5.9, with increments of 0.1. For each risk group, colonoscopy screening was optimized in terms of its start age (40-60 years), end age (70-85 years), and interval (1-20 years).

Main Outcome Measures: Quality-adjusted life years gained and costs compared to no screening, and threshold cost of polygenic testing.

Results: With current discriminatory performance, optimal screening ranged from no screening for those with an estimated RR<0.4 to 7-yearly colonoscopy at ages 45 to 75 for those with an estimated RR>2.2. This stratification reduced CRC screening and treatment costs with $141,000 per 1,000 40-year-olds while maintaining the same benefit as uniform screening. Consequently, at its current projected price of $200 per polygenic test, risk-stratified screening would be more expensive than uniform screening. Cost-savings could be achieved if the AUC value would increase to 0.65, or if the price per polygenic test would drop below $141.

Conclusions and relevance: Currently, CRC screening based on polygenic risk is unlikely to be cost-effective compared to uniform screening. This may change with an improved risk-stratification algorithm, or a lower price per polygenic test.

INTRODUCTION

Colorectal cancer (CRC) is the second most common cause of cancer death in the United States (US), with about 50,000 deaths expected in 2017.[1] Fortunately, screening and early treatment of adenomas and CRC can prevent CRC death.[2] Randomized controlled trials (RCTs) have shown that CRC mortality can be reduced by 15-30% with fecal occult blood testing, and by 40% with flexible sigmoidoscopy screening.[3-5] Colonoscopy screening is expected to achieve an even higher mortality reduction.[6]

Of US adults who were up-to-date with screening in 2012, two thirds were screened with colonoscopy.[7] As this is an invasive procedure, with the potential of serious complications,[8] it could be argued that those who would not have been diagnosed with CRC in the absence of screening (approximately 95% of the population) face unnecessary risks with colonoscopy screening. If screening could be targeted, then mortality could be reduced in those at increased CRC risk, while at the same time harms could be reduced in those at decreased CRC risk. To identify those at increased or decreased risk, exploratory studies have suggested the use of polygenic risk profiling.[9,10]

Polygenic risk prediction differs from commonly described hereditary CRC syndromes by its focus on multiple single nucleotide polymorphisms (SNPs) instead of a single genetic mutation. For individuals with identified inherited syndromes that are caused by a single genetic mutation, such as Lynch Syndrome, more intensive screening is already recommended.[11] In addition to the 2-5% of CRC cases that can be attributed to high penetrance genes causing these types of syndromes, an estimated 12-35% of CRC cases are caused by low penetrance genetic variants.[12] Thus far, genome-wide association studies (GWAS) have identified 37 common genetic variants that are associated with the risk for developing CRC.[13,14] Together, these variants explain approximately 14.5% of the familial relative risk in individuals of European descent.[14] As more genetic variants associated with CRC risk are detected this percentage might increase.[12]

The benefits of screening based on polygenic risk will depend on the discriminatory accuracy of risk algorithms in identifying those who will get CRC, as expressed by the area under the receiver operating characteristic curve (AUC). Based on the 37 currently identified common genetic variants associated with CRC risk, the AUC value was estimated at 0.6.[14] As more variants are discovered, this value may further increase. In this study, we explore the current and potential future benefits of risk-stratified screening based on common genetic variants. Additionally, we estimate at what cost per polygenic test risk-stratified screening would be cost-effective compared to current uniform screening.

METHODS

We used the Microsimulation Screening Analysis-Colon (MISCAN-Colon) model to simulate a population at average risk for CRC, that is willing to undergo polygenic testing and subse-

quent risk-stratified colonoscopy screening. We compared the results of such risk-stratified screening with those of current uniform screening, with colonoscopies at ages 50, 60, and 70.

MISCAN-Colon

MISCAN-Colon is a well-established microsimulation model for CRC developed at the Department of Public Health of the Erasmus University Medical Center (Rotterdam, the Netherlands).[15] The structure, underlying assumptions, and calibration of the model are described in the **Model appendix**. In brief, MISCAN-Colon simulates the life histories of a large population from age 40 until death. As each simulated person ages, one or more adenomas may develop. These adenomas can progress from small (≤5 mm), to medium (6-9 mm), to large size (≥10 mm). Some adenomas can develop into preclinical cancer, which may progress through stages I to IV. During each stage CRC may be diagnosed because of symptoms. Survival after clinical diagnosis is determined by the stage at diagnosis, the localization of the cancer, and the person's age.[16]

Screening will alter some simulated life histories through cancer prevented by adenoma detection and removal, or cancers detected at earlier stages resulting in more favorable survival estimates. However, screening can also result in serious complications and overdiagnosis and overtreatment of CRC (i.e. the detection and treatment of cancers that would not have been diagnosed without screening). By comparing a simulation of life histories with screening to a simulation of the same life histories without screening, MISCAN-Colon quantifies the effectiveness of screening, as well as the associated costs.

MISCAN-Colon was calibrated to the age-, stage-, and localization-specific incidence of CRC as observed in the Surveillance, Epidemiology, and End Results (SEER) Program before the introduction of screening (i.e. between 1975 and 1979),[17] and the age-specific prevalence and the multiplicity distribution of adenomas as observed in autopsy studies.[18-26] The preclinical duration of CRC and the adenoma dwell-time were calibrated to the rates of screen-detected and interval cancers observed in RCTs evaluating screening using guaiac fecal occult blood tests and a once-only sigmoidoscopy.[4,27-30] The model allows for a large variability in progression rates, enabling some adenomas to progress to CRC within a few years. More detailed information about MISCAN-Colon is available from the authors upon request.

Simulated population

We simulated a large cohort of 40-year-olds, with life expectancy as currently observed in the US, and followed them until death.[31] At model initiation, every simulated individual is assigned a background risk of developing adenomas. Given the development of an adenoma, the probability of progression to cancer does not differ between individuals.

Data and assumptions for polygenic testing

By testing for the presence of SNPs that are associated with CRC, a polygenic test can estimate someone's RR of developing CRC. A previously published method was used to generate the distribution of RR within an entire population, when assuming an AUC value of polygenic testing of 0.60, 0.65, 0.70, 0.75, and 0.80 (**Appendix Figure**).[32] We used an elliptical copula to ensure that the higher the AUC value, the higher the probability that someone's RR provides a good estimate of his/her background risk as simulated in MISCAN-Colon. For risk-stratified screening, RR distributions were split into 60 risk groups with RR varying from <0.1 to >5.9 using increments of 0.1.

Data and assumptions for screening and surveillance

We simulated scenarios without screening and with colonoscopy screening at intervals of 1, 2, 3, 5, 7, 10, 15, and 20 years. In addition, different start ages (40, 45, 50, 55, and 60) and different end ages (70, 75, 80, and 85) for screening were considered.

Test characteristics and complication rates of colonoscopy were based on literature. The specificity was assumed to be 86%, and the sensitivity increased from 75% for small adenomas (≤5 mm) to 85% for medium-sized adenomas (6-9 mm) and to 95% for large adenomas (≥10 mm) and CRC (**Table 1**).[33] Complications requiring a hospital admission or emergency department visit increased exponentially with age.[34,35]

Individuals with adenomas detected and removed at a screening were assumed to undergo colonoscopy surveillance according to the current guidelines.[36] Because surveillance is meant to follow individuals at an increased risk more closely, the recommended surveillance interval was shortened in situations where it would have been longer than the screening interval. We assumed that surveillance continued until 5 years after the end age of screening. Adherence to screening and surveillance colonoscopies was assumed to be 100%.

Data and assumptions for costs and utilities

The assumed loss in quality of life due to CRC screening was 0.002 QALYs per colonoscopy (1.5 days at 0.5 utility) and 0.0027-0.0055 QALYs per complication of colonoscopy (2-4 days at 0.5 utility) (**Table 1**). We also assumed that life years (LYs) with CRC care are of lower quality than those without CRC care.[37]

The cost-effectiveness analyses were conducted from a modified societal perspective. We included both direct medical costs as well as patient time costs. However, direct non-health costs and costs of informal care givers were not included.[38] The costs of colonoscopies were based on 2014 Medicare payment rates and copayments (**Table 1**). For each type of complication, the average payment by Centers for Medicare and Medicaid Services (CMS) was calculated using frequency data on hospitalizations for colonoscopy complications from Elizabeth Drye, MD, SM, and Craig Parzynski, MS, of Yale University (personal communication). Net costs of CRC care were obtained from an analysis of SEER-Medicare linked data[39] (personal communication, Robin Yabroff, PhD, and Martin Brown, PhD, both formerly of the

Table 1. Model inputs: Test characteristics, utility loss and costs of colonoscopy screening and treatment.

COLONOSCOPY TEST CHARACTERISTICS				
Specificity	86%[a]			
Sensitivity				
Small adenomas (≤5 mm)	75%[b]			
Medium-sized adenomas (6-9 mm)	85%[b]			
Large adenomas (≥10 mm)	95%[b]			
Colorectal cancer	95%[b]			
Reach	95% reaches the cecum; the reach of the remaining 5% is distributed uniformly over colon and rectum			
Complication rate for positive test				
Serious gastrointestinal event[c]	Age-specific[d]			
Other gastrointestinal event[e]	Age-specific[f]			
Cardiovascular event[g]	Age-specific[h]			
Mortality rate				
Positive test	0.0191 per 1,000[i]			
Negative test	0			
	UTILITY LOSS (QALYs)[j]			
Per colonoscopy	0.0020			
Per complication of colonoscopy				
Serious gastrointestinal event[c]	0.0055			
Other gastrointestinal event[e]	0.0027			
Cardiovascular event[g]	0.0048			
Per LY with CRC care[k,l]	Initial care	Continuing care	Terminal care *Death CRC*	Terminal care *Death other cause*
Stage I CRC	0.12	0.05	0.70	0.05
Stage II CRC	0.18	0.05	0.70	0.05
Stage III CRC	0.24	0.24	0.70	0.24
Stage IV CRC	0.70	0.70	0.70	0.70
	COSTS (2014 US$)[m]			
Polygenic test	200[n]			
Per colonoscopy				
without polypectomy/biopsy	1,422			
with polypectomy/biopsy	1,699			
Per complication of colonoscopy				
Serious gastrointestinal event[c]	11,142			
Other gastrointestinal event[e]	7,587			
Cardiovascular event[g]	8,453			

Table 1 (continued). Model inputs: Test characteristics, utility loss and costs of colonoscopy screening and treatment.

Per LY with CRC care[k]	Initial care	Continuing care	Terminal care *Death CRC*	Terminal care *Death other cause*
Stage I CRC	36,883	3,106	64,110	19,331
Stage II CRC	49,475	2,918	63,856	17,429
Stage III CRC	60,033	4,068	67,353	21,620
Stage IV CRC	78,124	12,274	88,749	50,122

QALY = quality-adjusted life year; LY = life year; CRC = colorectal cancer.

[a] We assumed that in 14% of all negative colonoscopies a non-adenomatous lesion was detected, resulting in a polypectomy or a biopsy, respectively.

[b] The sensitivity of colonoscopy for the detection of adenomas and CRC within the reach of the endoscope was obtained from a systematic review on miss rates observed in tandem colonoscopy studies.[33]

[c] Serious gastrointestinal events are perforations, gastrointestinal bleeding, or transfusions.

[d] Formula: $1/[\exp(9.27953 - 0.06105 \times \text{Age}) + 1] - 1/[\exp(10.78719 - 0.06105 \times \text{Age}) + 1]$

[e] Other gastrointestinal events are paralytic ileus, nausea and vomiting, dehydration, or abdominal pain.

[f] Formula: $1/[\exp(8.81404 - 0.05903 \times \text{Age}) + 1] - 1/[\exp(9.61197 - 0.05903 \times \text{Age}) + 1]$

[g] Cardiovascular events are myocardial infarction or angina, arrhythmias, congestive heart failure, cardiac or respiratory arrest, syncope, hypotension, or shock.

[h] Formula: $1/[\exp(9.09053 - 0.07056 \times \text{Age}) + 1] - 1/[\exp(9.38297 - 0.07056 \times \text{Age}) + 1]$

[i] Risk of dying from a colonoscopy at age 65 (Warren et al.[35], Gatto et al.[51] and Van Hees et al.[34])

[j] The loss of quality of life associated with a particular event.

[k] Care for CRC was divided in three clinically relevant phases: the initial, continuing, and terminal care phase. The initial care phase was defined as the first 12 months after diagnosis; the terminal care phase was defined as the final 12 months of life; the continuing care phase was defined as all months in between. In the terminal care phase, we distinguished between CRC patients dying from CRC and CRC patients dying from another cause. For patients surviving less than 24 months, the final 12 months were allocated to the terminal care phase and the remaining months were allocated to the initial care phase.

[l] Utility losses for LYs with initial care were derived from a study by Ness and colleagues.[37] For LYs with continuing care for stage I and II CRC, we assumed a utility loss of 0.05 QALYs; for LYs with continuing care for stage III and IV CRC, we assumed the corresponding utility losses for LYs with initial care. For LYs with terminal care for CRC, we assumed the utility loss for LYs with initial care for stage IV CRC. For LYs with terminal care for another cause, we assumed the corresponding utility losses for LYs with continuing care.

[m] Costs include copayments and patient time costs (i.e. the opportunity costs of spending time on screening or being treated for a complication or CRC), but do not include travel costs, costs of lost productivity, and unrelated health care and non-health care costs in added years of life. We assumed that the value of patient time was equal to the median wage rate in 2014: $17.09 per hour.[52] We assumed that colonoscopies used up 36 hours, serious gastrointestinal complications 192 hours, other gastrointestinal complications 96 hours and cardiovascular complications 120 hours of patient time. Patient time costs associated with CRC care were provided by Yabroff (personal communication), and were calculated using the methodology described in a study by Yabroff and colleagues.[53]

[n] Polygenic testing costs were based on a currently available polygenic test.[41]

National Cancer Institute). Patient time costs and copayments were added to all of these estimates, which were then updated to 2014 US dollars using the Consumer Price Index.[40] For polygenic testing we assumed a cost of $200, based on the price of a currently available polygenic test.[41]

Analyses and outcomes

For every RR group we simulated all screening strategies to quantify their QALYs and costs as compared to no screening. For uniform screening, we simply summed the results of colonoscopy screening at ages 50, 60, and 70 over all RR groups. For risk-stratified screening, we optimized screening for each RR group. To this end, we first determined a list of efficient screening strategies by excluding all dominated screening strategies, i.e. those strategies that were more costly and less effective than other strategies. We then ranked the remaining strategies based on the number of QALYs gained and calculated their incremental cost-effectiveness ratio (ICER) compared to the next less effective, efficient strategy. For every RR group, we then defined the optimal screening strategy as the strategy with an ICER just below the willingness-to-pay (WTP) threshold of $50,000 per QALY gained. Finally, we summed the results of screening the RR groups with their optimal strategies to obtain population-level outcomes for risk-stratified screening. The WTP threshold was also adjusted to a level at which the QALYs gained of risk-stratified screening were equal to those of uniform screening. In this way, we could determine the potential cost-savings of replacing current uniform screening with risk-stratified screening. We applied the conventional 3% annual discount rate for both costs and effects.

Threshold analyses

As long as polygenic testing is not yet available to the general population, the cost of polygenic testing is uncertain. In the basecase analysis we assumed a value of $200 per test, based on the current price of a commercially available polygenic test.[41] In threshold analyses, we adjusted the WTP threshold such that the health benefits of risk-stratified screening were equal to those of uniform screening. This enabled us to estimate the maximum price per polygenic test for which risk-stratified screening would be more cost-effective than uniform screening (i.e., the threshold cost of polygenic testing).

Sensitivity analyses

In one-way sensitivity analyses, we assumed:

1. a simplified version of risk-stratified screening considering only three risk groups: "low", "moderate" and "high", which were approximately equal in size, and
2. offering at least one screening colonoscopy to every risk group.

RESULTS

Compared to no screening, uniform screening (i.e. screening all individuals at ages 50, 60, and 70) yielded 73 LYs and 85 QALYs per 1,000 40-year-olds, at a total cost of $1,626,000 (**Table 2**). With current discriminatory performance, optimal screening strategies ranged from no screening for those with an estimated RR below 0.4 to 7-yearly colonoscopy screening at ages 45-75 for those with an estimated RR above 2.2 (**Figure 1**). A risk-stratified

screening program including those strategies resulted in 68 LYs and 80 QALYs gained at a discounted CRC-related cost of $1,159,000. When costs of polygenic testing were included, total costs equaled $1,359,000.

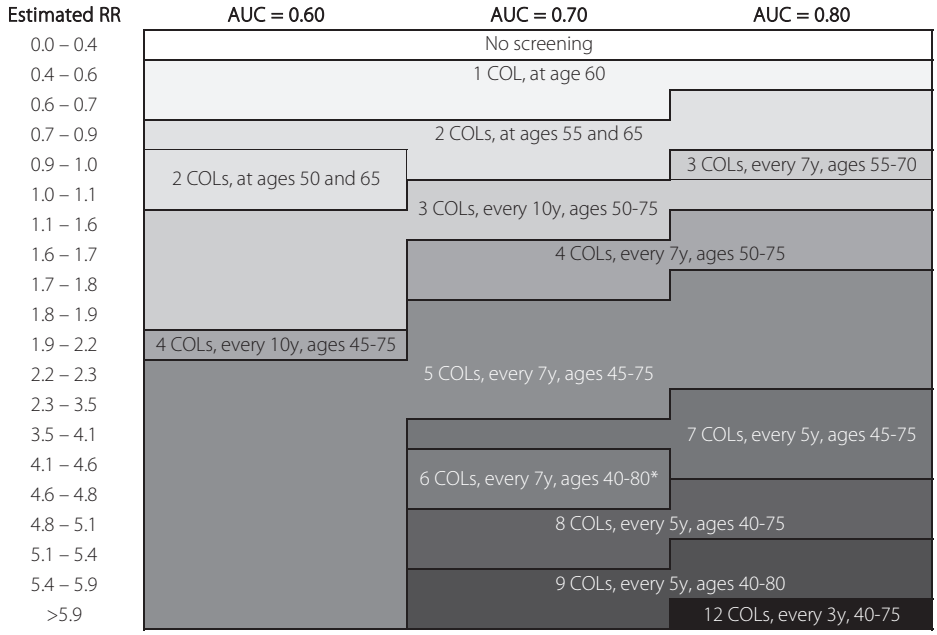


Figure 1. Cost-effective strategies (under a willingness-to-pay threshold of $50,000 per QALY) by relative risk as estimated by a polygenic test with an AUC value of 0.60, 0.70, and 0.80. For every strategy, the number of lifetime colonoscopies, screening interval and age range of screening is given (i.e. "3 COLs, every 10y, ages 50-75" refers to 3 lifetime colonoscopies with an interval of 10 years in individuals aged 50-75).
RR = relative risk; AUC = area under the receiver operating characteristic curve; COLs = colonoscopies; y = years.
* Individuals with an estimated RR of 4.1-4.8 are offered fewer lifetime screens than those with an estimated RR of 3.5-4.1, but the age range in which they are offered screening is broader.

Threshold cost

At current discriminatory performance, risk-stratified screening would reduce CRC-related costs by $141,000 per 1,000 40-year-olds while maintaining the same benefits as uniform screening. This suggests that risk-stratified screening would be cost-effective if the costs of polygenic testing would not exceed $141 per person (**Figure 2**). This maximum allowable cost increased with increasing discriminatory performance, from $291 at an AUC of 0.65, to $1,112 at an AUC of 0.80.

Sensitivity analyses

Restricting the number of risk groups to a maximum of 3 decreased the potential benefit of risk-stratified screening and thus the threshold cost of the polygenic test (**Figure 2**). At AUC = 0.60, the threshold was even negative, suggesting that risk-stratified screening would be

Table 2. Effects and costs per 1,000 40-year-old individuals for no screening, uniform screening and risk-stratified screening for a willingness to pay of $50,000 per QALY gained.

	Colonoscopies		CRC cases		CRC deaths		Life years[a]		QALYs[a]		Costs, USD (*1,000)[a,b]							
											Polygenic testing		CRC screening[c]		Cancer diagnosis and treatment		Total	
No screening	67	(ref)	67	(ref)	28	(ref)	22,940	(ref)	22,908	(ref)	0	(ref)	4	(ref)	2,476	(ref)	2,480	(ref)
Uniform screening	3,247	(+3,180)	30	(-37)	8	(-20)	23,014	(+73)	22,993	(+85)	0	(--)	2,809	+2,805	1,299	-1,178	4,107	(+1,626)
Risk-stratified screening																		
AUC = 0.60 (base case)	2,715	(+2,648)	30	(-36)	8	(-19)	23,009	(+68)	22,987	(+80)	200	(+200)	2,270	(+2,266)	1,370	(-1,107)	3,840	(+1,359)
AUC = 0.65	2,590	(+2,523)	30	(-36)	9	(-19)	23,009	(+68)	22,988	(+80)	200	(+200)	2,179	(+2,175)	1,360	(-1,117)	3,740	(+1,259)
AUC = 0.70	2,432	(+2,365)	30	(-36)	9	(-19)	23,011	(+70)	22,990	(+83)	200	(+200)	2,083	(+2,079)	1,327	(-1,150)	3,611	(+1,130)
AUC = 0.75	2,339	(+2,272)	29	(-37)	8	(-19)	23,013	(+73)	22,994	(+86)	200	(+200)	2,033	(+2,029)	1,276	(-1,201)	3,509	(+1,028)
AUC = 0.80	2,296	(+2,230)	28	(-39)	8	(-20)	23,017	(+7 6)	22,998	(+91)	200	(+200)	2,028	(+2,024)	1,203	(-1,274)	3,432	(+951)

AUC = area under the receiver operating characteristic curve; CRC = colorectal cancer; QALYs = quality-adjusted life years; ref = reference value; (n) = increase/decrease compared to no screening.

[a] (Quality-adjusted) life years and costs were discounted at an annual rate of 3%.

[b] Costs are in 2014 US Dollars.

[c] Includes costs of screening colonoscopies, surveillance colonoscopies, and colonoscopy complications.

-inefficient irrespective of the price per polygenic test. Compared to the base-case analysis, offering every risk group at least one colonoscopy resulted in similar current benefits, but did reduce potential future benefits.

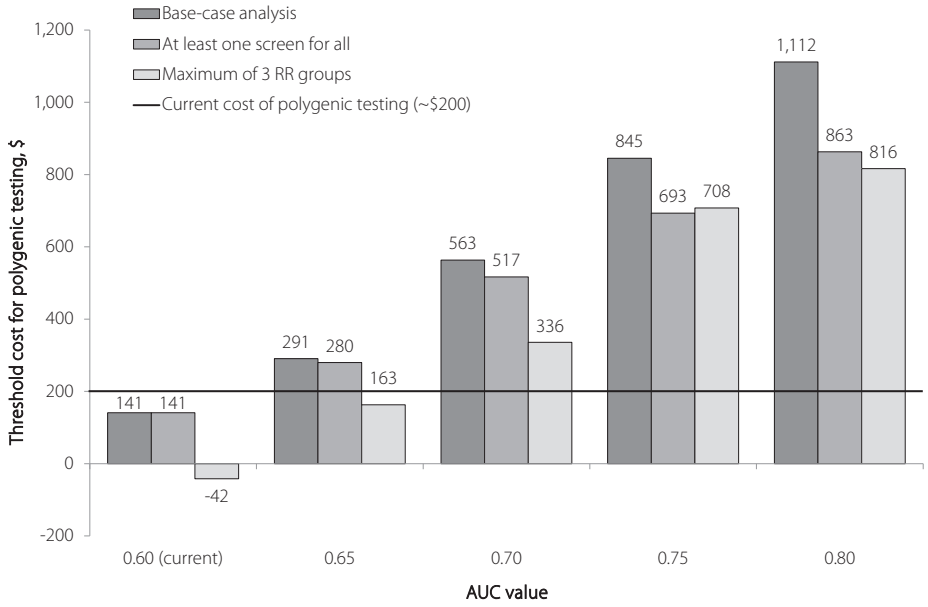


Figure 2. Estimated cost per test for which risk-stratified screening would be equally cost-effective as uniform screening at ages 50, 60, and 70.
AUC = area under the receiver operating characteristic curve.

DISCUSSION

This study shows that, under current discriminatory performance of polygenic testing (i.e. AUC = 0.60), the health benefits of risk-stratified screening are modest at most. At a current estimated price of $200 per polygenic test, risk-stratified screening is not cost-effective compared to uniform screening. This could change if the costs of polygenic testing would drop below $141 per person, or if the AUC value of polygenic testing would increase beyond 0.65.

In theory, when using the same amount of resources, risk-stratified screening based on any algorithm with positive discriminatory performance is more efficient than uniform screening. The reason for this is that available resources, in this case colonoscopies, are not distributed equally, but are allocated based on their expected yield. Indeed, risk-stratified screening was consistently more efficient than uniform screening in our model results. However, in practice, risk-stratified screening can easily lead to a less efficient outcome. If the increased complexity or potential problematic acceptability of a risk-stratified screening program leads to a reduction in colonoscopy uptake, then benefits of risk-stratification can

be offset. Complexity and acceptability issues could partly be solved by including fewer risk categories and offering everyone at least one lifetime screen. Sensitivity analyses showed that especially including fewer risk categories could decrease the potential benefits of risk-stratified screening.

This study represents an early exploration of the potential benefit of risk-stratified screening based on common genetic variants. We assumed full adherence with both genetic testing and subsequent recommended screening. In reality, people may refrain from polygenic testing due to various well-grounded reasons,[42] and if tested, may not be screened according to their optimal screening strategy. In a recent US survey among people with an intermediate familial CRC risk, three-fourths of participants said that they would probably (47%) or definitely (27%) have SNP testing to estimate their CRC risk.[43] Actual population uptake of polygenic testing will depend on the implementation of the program and the information provided to the public. Although individuals with a family history of CRC are more adherent with CRC screening guidelines,[44] increased awareness does not necessarily lead to increased adherence. One US study showed that knowledge of elevated CRC risk increased screening participation from 50 to 67% in whites but, surprisingly, decreased screening participation from 54 to 33% in non-whites.[45]

By varying the AUC value and other uncertain parameters, we considered a broad spectrum of potential current and future applications of risk-stratified screening. Nevertheless, this study has some limitations. First of all, we did not assume any disutility for having a polygenic test and for knowing your polygenic risk profile. Even though knowing that you have an increased CRC risk can be burdensome, it also enables people to get more intensive colonoscopy screening that could improve early detection of cancer. Moreover, the majority will be reassured by a relatively low CRC risk. Second, we presumed that our ability to predict risk for CRC is equally robust across the entire spectrum of risk (i.e. for those at decreased as well as increased risk), which remains to be demonstrated empirically.

We showed that risk-stratified colonoscopy screening may become clinically relevant as more genetic variants associated with CRC risk are being identified. Although AUC values above 0.75 may be hypothetically high, it has been estimated that from the total number of SNPs associated with CRC risk, less than 10% has currently been identified.[12] Another way of increasing the AUC value of risk-stratification algorithms is by including other risk factors, such as gender and family history of CRC[46], and potentially also lifestyle. While physical activity has shown to have a protective effect for CRC (RR of 0.81), alcohol consumption, diabetes, red meat and processed meat consumption, obesity and smoking have all been associated with an increased risk for CRC (RRs ranging from 1.56 to 1.16).[47] Although screening based on lifestyle may seem controversial and hard to achieve in practice, it is already being implemented with the US Preventive Services Task Force current recommendation of lung cancer screening in heavy smokers only.[48]

A comparable study concluded that risk-stratification based on SNPs could also improve the efficiency of the stool-based screening program that is currently in place in the United Kingdom.[14] As compared to uniform screening, they estimated that personalized screening

would reduce the number of men and women being eligible for screening by 16% and 17% respectively, at a cost of 10% and 8% fewer screen-detected cases. Another study showed similar results for breast cancer and prostate cancer screening.[49] If, in the future, a single polygenic test would be available to estimate a person's risk for multiple cancers or other genetic diseases simultaneously, then benefits of polygenic testing could be achieved more easily.

In conclusion, with current discriminatory performance of polygenic testing, the benefits of risk-stratified screening based on polygenic risk are modest at most. Given the additional costs of polygenic testing, risk-stratified screening is not (yet) a cost-effective alternative to uniform screening. However, this might change because of future developments. If more variants associated with CRC risk would be identified, then the discriminatory performance of polygenic testing would increase, and so would the benefits of risk-stratified screening. Risk-stratification algorithms could also be enhanced by including other risk factors, such as lifestyle. Finally, the costs of polygenic testing could be lower when offered on a population level because of economies of scale, and the costs allocated to CRC risk estimation could be lower than those of polygenic testing itself if its results would also be used for other purposes.

ACKNOWLEDGMENTS

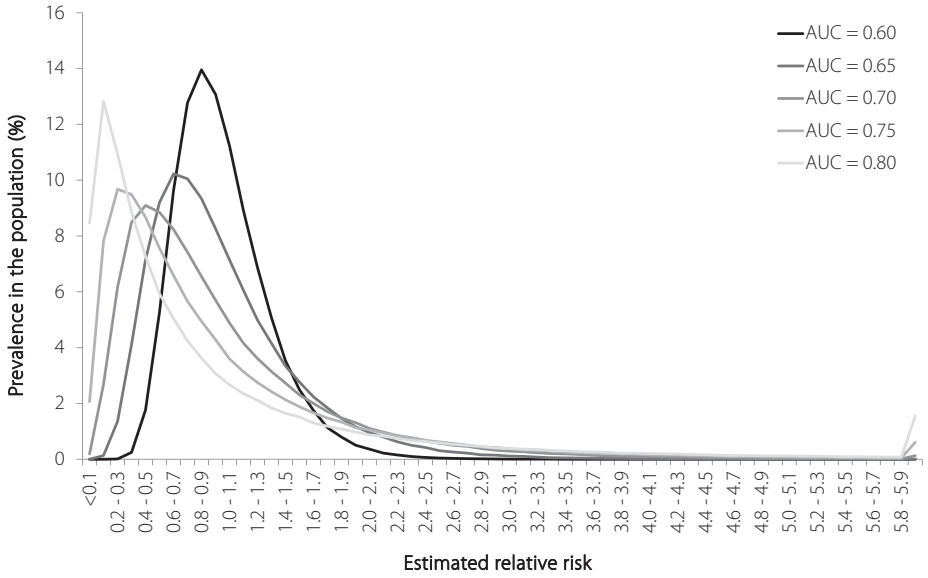
This work resulted from a collaboration between the Cancer Intervention and Surveillance Modeling Network (CISNET) and the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. The authors would like to thank Elizabeth Drye, MD, SM, and Craig Parzynski, MS, of Yale University, who provided us with frequency data on hospitalizations for colonoscopy complications.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. CA Cancer J Clin 2017;67:7-30.
2. Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 2012;366:687-96.
3. Hewitson P, Glasziou P, Watson E, et al. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. Am J Gastroenterol 2008;103:1541-9.
4. Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. Lancet 2010;375:1624-33.
5. Segnan N, Armaroli P, Bonelli L, et al. Once-only sigmoidoscopy in colorectal cancer screening: follow-up findings of the Italian Randomized Controlled Trial--SCORE. J Natl Cancer Inst 2011;103:1310-22.
6. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, et al. Evaluating test strategies for colorectal cancer screening: a decision analysis for the U.S. Preventive Services Task Force. Ann Intern Med 2008;149:659-69.
7. Centers for Disease Control and Prevention. Vital signs: colorectal cancer screening test use--United States, 2012. MMWR Morb Mortal Wkly Rep 2013;62:881-8.
8. Whitlock EP, Lin JS, Liles E, et al. Screening for colorectal cancer: a targeted, updated systematic review for the U.S. Preventive Services Task Force. Ann Intern Med 2008;149:638-58.
9. Khoury MJ, Janssens AC, Ransohoff DF. How can polygenic inheritance be used in population screening for common diseases? Genet Med 2013;15:437-43.
10. Hawken SJ, Greenwood CM, Hudson TJ, et al. The utility and predictive value of combinations of low penetrance genes for screening and risk prediction of colorectal cancer. Hum Genet 2010;128:89-101.
11. Lindor NM, Petersen GM, Hadley DW, et al. Recommendations for the care of individuals with an inherited predisposition to Lynch syndrome: a systematic review. JAMA 2006;296:1507-17.
12. Jiao S, Peters U, Berndt S, et al. Estimating the heritability of colorectal cancer. Hum Mol Genet 2014;23:3898-905.
13. Montazeri Z, Theodoratou E, Nyiraneza C, et al. Systematic meta-analyses and field synopsis of genetic association studies in colorectal adenomas. Int J Epidemiol 2016;45:186-205.
14. Frampton MJ, Law P, Litchfield K, et al. Implications of polygenic risk for personalised colorectal cancer screening. Ann Oncol 2016;27:429-34.
15. Loeve F, Boer R, van Oortmarssen GJ, et al. The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening. Comput Biomed Res 1999;32:13-33.
16. Rutter CM, Johnson EA, Feuer EJ, et al. Secular trends in colon and rectal cancer relative survival. J Natl Cancer Inst 2013;105:1806-13.
17. Surveillance, Epidemiology, and End Results (SEER) Program SEER*Stat Database: Incidence - SEER 9 Regs Limited-Use, Nov 2002 Sub (1973-2000). National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2003, based on the November 2002 submission, 2003. Accessed at: www.seer.cancer.gov.
18. Arminski TC, McLean DW. Incidence and Distribution of Adenomatous Polyps of the Colon and Rectum Based on 1,000 Autopsy Examinations. Dis Colon Rectum 1964;7:249-61.
19. Chapman I. Adenomatous polypi of large intestine: incidence and distribution. Ann Surg 1963;157:223-6.
20. Bombi JA. Polyps of the colon in Barcelona, Spain. An autopsy study. Cancer 1988;61:1472-6.
21. Clark JC, Collan Y, Eide TJ, et al. Prevalence of polyps in an autopsy series from areas with varying incidence of large-bowel cancer. Int J Cancer 1985;36:179-86.

22. Jass JR, Young PJ, Robinson EM. Predictors of presence, multiplicity, size and dysplasia of colorectal adenomas. A necropsy study in New Zealand. Gut 1992;33:1508-14.
23. Johannsen LG, Momsen O, Jacobsen NO. Polyps of the large intestine in Aarhus, Denmark. An autopsy study. Scand J Gastroenterol 1989;24:799-806.
24. Rickert RR, Auerbach O, Garfinkel L, et al. Adenomatous lesions of the large bowel: an autopsy survey. Cancer 1979;43:1847-57.
25. Vatn MH, Stalsberg H. The prevalence of polyps of the large intestine in Oslo: an autopsy study. Cancer 1982;49:819-25.
26. Williams AR, Balasooriya BA, Day DW. Polyps and cancer of the large bowel: a necropsy study in Liverpool. Gut 1982;23:835-42.
27. Hardcastle JD, Chamberlain JO, Robinson MH, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet 1996;348:1472-7.
28. Jorgensen OD, Kronborg O, Fenger C. A randomised study of screening for colorectal cancer using faecal occult blood testing: results after 13 years and seven biennial screening rounds. Gut 2002;50:29-32.
29. Lansdorp-Vogelaar I, van Ballegooijen M, Boer R, et al. A novel hypothesis on the sensitivity of the fecal occult blood test: Results of a joint analysis of 3 randomized controlled trials. Cancer 2009;115:2410-9.
30. Mandel JS, Church TR, Ederer F, et al. Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. J Natl Cancer Inst 1999;91:434-7.
31. Arias E. United States Life Tables, 2010. National Vital Statistics Reports. Volume 63: Centers for Disease Control and Prevention (CDC), 2014.
32. Janssens AC, Aulchenko YS, Elefante S, et al. Predictive testing for complex diseases using multiple genes: fact or fiction? Genet Med 2006;8:395-400.
33. van Rijn JC, Reitsma JB, Stoker J, et al. Polyp miss rate determined by tandem colonoscopy: a systematic review. Am J Gastroenterol 2006;101:343-50.
34. van Hees F, Zauber AG, Klabunde CN, et al. The appropriateness of more intensive colonoscopy screening than recommended in medicare beneficiaries: a modeling study. JAMA Intern Med 2014;174:1568-76.
35. Warren JL, Klabunde CN, Mariotto AB, et al. Adverse events after outpatient colonoscopy in the Medicare population. Ann Intern Med 2009;150:849-57, W152.
36. Lieberman DA, Rex DK, Winawer SJ, et al. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 2012;143:844-57.
37. Ness RM, Holmes AM, Klein R, et al. Utility valuations for outcome states of colorectal cancer. Am J Gastroenterol 1999;94:1650-7.
38. Gold MR, Siegel JE, Russell LB, et al. Cost-effectiveness in Health and Medicine. New York: Oxford University Press, 1996.
39. Yabroff KR, Lamont EB, Mariotto A, et al. Cost of care for elderly cancer patients in the United States. J Natl Cancer Inst 2008;100:630-41.
40. The United States Department of Labor. Bureau of Labor Statistics. Consumer Price Index.
41. Find out what your DNA says about you and your family. http://www.23andme.com; Accessed 09-02-2017.
42. Otlowski M, Taylor S, Bombard Y. Genetic discrimination: international perspectives. Annu Rev Genomics Hum Genet 2012;13:433-54.
43. Anderson AE, Flores KG, Boonyasiriwat W, et al. Interest and informational preferences regarding genomic testing for modest increases in colorectal cancer risk. Public Health Genomics 2014;17:48-60.

44. James TM, Greiner KA, Ellerbeck EF, et al. Disparities in colorectal cancer screening: a guideline-based analysis of adherence. Ethn Dis 2006;16:228-33.
45. Myers RE, Ruth K, Manne SL, et al. Effects of genetic and environmental risk assessment feedback on colorectal cancer screening adherence. J Behav Med 2015;38:777-86.
46. Dunlop MG, Tenesa A, Farrington SM, et al. Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42,103 individuals. Gut 2013;62:871-81.
47. Huxley RR, Ansary-Moghaddam A, Clifton P, et al. The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence. Int J Cancer 2009;125:171-80.
48. Moyer VA, U. S. Preventive Services Task Force. Screening for cervical cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2012;156:880-91, W312.
49. Pashayan N, Duffy SW, Chowdhury S, et al. Polygenic susceptibility to prostate and breast cancer: implications for personalised screening. Br J Cancer 2011;104:1656-63.



Appendix Figure 1. Distribution of relative risk that is assumed to be revealed by a polygenic test having an AUC value of 0.60 (i.e., current discriminatory performance) and of 0.65, …, 0.80 (i.e., potential future discriminatory performance).

AUC = area under the receiver operating characteristic curve.

Chapter 11

General discussion

This chapter starts with a summary of the main findings from the studies that are presented in this thesis. We will then interpret those findings, and elaborate on their implications for cervical cancer and colorectal cancer screening. The chapter ends with conclusions and recommendations.

MAIN FINDINGS

This section provides answers to the research questions that were raised in the introduction of this thesis.

Methodological issues in cost-effectiveness analyses

In general, cost-effectiveness analyses (CEAs) are used to compare the effectiveness and cost effectiveness of different screening strategies. The incremental cost-effectiveness ratio (ICER) is a commonly used measure to judge the cost-effectiveness of cancer screening strategies. For all efficient strategies, the ICER equals the additional amount of money that needs to be paid to save one additional (healthy) life year, compared to the next less effective strategy. The estimated ICER of one strategy therefore depends on the estimated costs and health benefits of other included strategies. The strategy with the highest ICER within the willingness-to-pay threshold is deemed optimal.

To what extent do cost-effectiveness analyses of cervical cancer screening omit relevant strategies, and how does this affect their conclusions?

In the review presented in **Chapter 2**, we identified 30 CEAs of cervical cancer screening using HPV testing that were published between January 1995 and September 2013. We compared the set of included strategies and their estimated ICERs. Eighteen studies were likely to have omitted one or more strategies in the CEA-relevant portion of the efficiency frontier, i.e. with ICERs around the willingness-to-pay threshold. In all but three of those studies, a longer interval than the one found to be optimal was not included. It should be acknowledged that even though strategies with a longer interval may not be considered by policy makers, their inclusion is necessary to reliably estimate ICERs of strategies that are considered. The underestimation of ICERs can lead to the acceptance and implementation of strategies that are not cost-effective. Readers of CEAs can recognize this type of methodological flaw by the apparent 'kinks' in efficiency frontiers.

Reducing harms

Although cancer screening programs aim to reduce morbidity and mortality, they always come with burden and harm for some of its participants. In general, for screening tests, there is a trade-off between high sensitivity and high specificity. As many cervical lesions tend to regress naturally, cytology screening is characterized by a relatively low specificity. Unnecessary follow-up testing and potential treatment of regressive cervical intraepithelial

neoplasia (CIN) may involve substantial harms, which are likely to increase when cytology is replaced with HPV-based screening.

What are the potential harms of primary HPV screening in over-screened women?

As HPV-based screening offers better protection against (precursors of) cervical cancer than cytology screening,[1] it is implemented in both Australia and the Netherlands in 2017,[2,3] and is considered by several other developed countries. The disadvantage of HPV-based screening, however, is that most HPV infections clear naturally and are therefore preferably not detected by screening. Especially at young age, a lot of women who are infected with HPV will never develop neoplasia, let alone cancer.

In **Chapter 3**, we estimated the potential harms of HPV-based screening for over-screened women, who are screened more frequently and from a younger age than recommended by international guidelines. We found that for women screened every 5 years from age 30, switching to HPV-based screening resulted in 32 fewer cervical cancer deaths at a cost of ~9,000 more positive primary tests. For those screened annually from age 20, this ratio was much worse, with only 6-7 fewer deaths at a cost of almost 77,000 more positive primary tests. Since with annual or biennial cytology-based screening, clinically relevant lesions are rarely missed, a switch to HPV-based testing will merely reduce cancer incidence and mortality in women who are screened this frequently. However, HPV testing will lead to many more positive tests and (false-positive) referrals to colposcopy. Altogether, for over-screened women, switching to HPV-based screening was found to be associated with a net loss in health.

What is the impact of cervical screening on preterm birth? Is it of such importance that cervical screening at reproductive age should depend on a woman's childwish?

Loop electrosurgical excision procedure (LEEP) is the most commonly used method to remove CIN.[4] Although the procedure is effective in removing CIN, women treated with LEEP have a 39% higher risk of subsequent preterm birth than women with untreated CIN.[5] Besides reduced survival probabilities and potential disabilities for the child itself, a preterm birth results in substantial costs and emotional distress for parents.[6] To our knowledge, this harm has not been taken into account in (cost-)effectiveness analyses of cervical cancer screening. Therefore, **Chapter 4** of this thesis provided insight into the impact of cervical cancer screening on preterm birth risk and subsequent infant morbidity and mortality.

Using US birth rates and baseline risk of preterm birth, we showed that initiating 3-yearly screening at age 21 instead of age 30 yields very limited health gain in women of reproductive age (i.e. ~300 life years gained per 100,000 screened women). Moreover, even without considering the negative impact of increased preterm birth risk, it resulted in a net loss in quality-adjusted life years. Lowering the start age of 3-yearly screening from 30 to 21 years did significantly increase preterm birth rates with 223 per 100,000 screened women, leading to 20 more infants with (severe) morbidity and 6 more stillborn infants.

Given that health benefits of cervical cancer screening below age 30 are limited (or even negative), but preterm birth risk increases substantially, delaying the start age of screening to age 30 should be considered, especially in women with a (future) childwish.

Increasing benefits

Benefits of screening can be enhanced by switching to a screening test that provides better protection against invasive cancer. Given that incidence and mortality is overrepresented in non-attendees, perhaps the most effective way of increasing benefits is by increasing the adherence with screening. Introducing an alternative screening test, like the self-sampling HPV test for cervical cancer screening and the multitarget stool DNA test for colorectal cancer screening, can entice individuals who would otherwise not have participated in screening, to participate in screening.

Does cervical cancer incidence after a normal cytological sample differ between SurePath, ThinPrep and conventional cytology?

Over the past decade, liquid-based cytology (LBC) tests SurePath and ThinPrep have gradually replaced conventional cytology in many countries with organized cervical cancer screening programs, such as the UK, the Netherlands and Denmark. As compared to conventional cytology, the advantages of using LBC are that it facilitates co-testing (i.e. the LBC sample can also be used for HPV testing) and that it results in fewer samples of unsatisfactory quality. Two systematic reviews have shown similar sensitivity for LBC and conventional cytology,[7,8] but did not make a distinction between SurePath and ThinPrep. In an earlier study, we have shown that compared to the use of conventional cytology, the use of SurePath results in increased CIN 2+ detection rates, while the use of ThinPrep does not.[9] The remaining question was whether those additionally detected CIN 2+ lesions would all have regressed in the absence of screening, or whether their detection prevented development to cervical cancer.

To answer this question, we have calculated the cumulative cancer incidence after a normal SurePath, ThinPrep and conventional cytology screening sample taken within the Dutch cervical cancer screening program (**Chapter 5**). Using a Cox regression analysis, we showed that the 6-year risk of cervical cancer after a normal screening sample was 19% lower for SurePath than for conventional cytology. No significant difference was found between ThinPrep and conventional cytology. The assumed similarity of test characteristics between LBC and conventional cytology, and between different types of LBC tests, should therefore be reconsidered.

When do the harms of offering HPV self-sampling to non-attendees of organized primary HPV screening outweigh its benefits?

The aim of offering self-sampling to non-attendees of cervical cancer screening is to entice women who would otherwise not be screened, to participate in screening. However, it is possible that women who would otherwise be screened by a clinician, will switch from

getting the office-based test to using the self-sampling kit. If the test characteristics of the self-sampling HPV test are comparable to those of the office-based HPV test, this will not be problematic. It will in fact be cost-saving, given that self-sampling does not involve smear-taking costs by a clinician. However, the use of self-sampled material instead of material collected by a clinician could lead to a lower sensitivity and/or specificity of the HPV test.

In **Chapter 6**, we investigated the conditions under which offering self-sampling to non-attendees would be (cost-)effective. We found that self-sampling is expected to gain health effects if the relative CIN 2+ sensitivity is at least 0.95, previously unscreened women are recruited, and the total attendance increases with at least 6 percentage points. If these requirements are not met, women switching from office-based sampling to self-sampling may decrease the total effectiveness of the program.

Is colorectal cancer screening with the multitarget stool DNA test a cost-effective alternative for the Medicare population, and if not, under what conditions will it be?

The recently developed multitarget stool DNA (mtSDNA) test has been included in the 2016 colorectal cancer screening guidelines of the United States Preventive Services Task Force (USPSTF).[10] Three-yearly screening with this test is now covered by the Centers for Medicare and Medicaid Services (CMS). In **Chapter 7**, we have estimated whether this screening strategy is a cost-effective alternative to other reimbursed strategies for the Medicare population, and if not, at what cost, adherence or frequency it would be.

As compared to no screening, screening with the mtSDNA test reduces the incidence and mortality from colorectal cancer. However, at $493 per test, the total costs are much higher than for other stool-based tests ($22 per FIT and $4 per FOBT). Therefore, 3-yearly mtSDNA testing is not a cost-effective alternative to other reimbursed strategies. Even if adherence would increase with 30%, this conclusion did not change. The reimbursement rate has to be below $33 for 3-yearly mtSDNA testing to be cost-effective. Although per-test reimbursement rates of $58 and $47 could be supported with annual or biennial mtSDNA testing, these rates are still way lower than the current reimbursement rate of $493.

Reducing harms and increasing benefits: Opportunities for risk-based screening

Personalized screening provides the opportunity to reduce harms of screening in individuals at low risk for the disease, while increasing benefits in those at higher risk. Possibilities for personalization include cervical cancer screening based on HPV vaccination status and colorectal cancer screening based on genetic risk factors.

At what level of herd immunity can uniform cervical cancer screening be adjusted to the risk level in vaccinated women?

It is expected that HPV-vaccinated women will be at lower risk for developing cervical cancer than unvaccinated women. Using microsimulation modeling, we determined the optimal screening strategy for both a pre-vaccination, and a fully vaccinated cohort (**Chap-**

ter 8). Under a willingness-to-pay threshold of €50,000 per quality-adjusted life year, primary HPV screening with cytology triage was found to be most cost-effective for both cohorts. Whereas for the pre-vaccination cohort, the optimal strategy included 8 lifetime screens, for the fully vaccinated cohort, it included only 3.

Due to organizational or ethical matters, it may not be possible to offer screening based on HPV vaccination status. If a population will be offered uniform screening, then the intensity of such screening is preferably reduced as vaccination coverage, and thereby herd immunity, increases. Although it would be most cost-effective to reduce the intensity gradually, it may not be feasible to adjust the screening program every few years. When the screening program will only be adjusted once, then one might consider doing so when herd immunity reaches ~50%, as from then on it is more cost-effective to offer unvaccinated women 3 lifetime screens instead of 8. This result merely changed when the background cervical cancer risk was increased in unvaccinated women, or when the quadrivalent instead of the bivalent vaccine was assumed to be offered. Therefore, if a cervical cancer screening program cannot be tailored to HPV-vaccination status, and it can only be adjusted once, then one should wait with adjusting the screening frequency to the risk level in vaccinated women until unvaccinated women benefit from a herd immunity level of at least 50%.

Should colorectal cancer screening for people with a positive family history vary by age?

It is well-known that individuals with a family history of colorectal cancer are at an increased risk for developing the disease themselves. Therefore, several screening guidelines recommend more intensive screening for individuals with one or more affected first-degree relatives (FDRs) than for people in the general population.[11] Although for a given level of family history, the relative risk of developing colorectal cancer as compared to the general population decreases with age, the current guidelines do not let the recommended interval of screening vary by age. To investigate whether such a refinement would be worthwhile, in **Chapter 9** we optimized colonoscopy screening for individuals with 1, 2, 3, or ≥4 affected FDRs, allowing different screening intervals at different ages.

Just as in an earlier modeling study,[12] we found that it is cost-effective to offer more intensive screening to individuals with more affected FDRs. In addition, we found that for individuals with a constant level of family history, the screening interval can be lengthened after several subsequent colonoscopies. For individuals with one affected FDR, comprising ~92% of those with a positive family history, the optimal screening interval rapidly increased from 3 years at age 40 to 7 years at age 55. Individuals with 2 or 3 affected FDRs should start screening earlier and lengthen their screening interval more gradually, from 2-3 years at age 35 to 7 years at age 70. These results show that individuals with a constant level of family history can safely lengthen their screening interval, provided that they have had several subsequent negative colonoscopies.

What is the potential benefit of risk-stratified colorectal cancer screening based on common genetic variants?

Whereas individuals with well-defined cancer syndromes have one rare genetic mutation that increases their cancer risk substantially, the presence of multiple more common genetic variants can also increase someone's cancer risk significantly. Until now, relatively few of these so-called single nucleotide polymorphisms (SNPs) associated with colorectal cancer risk have been discovered. Consequently, currently available polygenic tests have limited discriminative power, with an area under the ROC curve (AUC) of approximately 0.60. In **Chapter 10**, we explored the current and potential future benefit of risk-based screening using polygenic information.

We showed that under current discriminatory performance of polygenic testing, the health benefits of risk-stratified screening are modest at most. At a current estimated price of $200 per polygenic test, risk-stratified screening is not cost-effective compared to uniform screening. This could change if the costs of polygenic testing would drop below $141 per person, or if the AUC value of polygenic testing would increase beyond 0.65.

INTERPRETATION OF FINDINGS

The studies presented in this thesis all aim at informing policy makers on the possible opportunities to improve the harm-benefit ratio of cervical cancer or colorectal cancer screening. While some results, like those presented in **Chapter 5**, could be of direct use; others come from more exploratory analyses that may not directly result in policy applications but rather provide guidance for future developments and research. In the paragraphs below, we describe the usefulness of the outcomes for policy makers.

Improving adherence with cervical cancer screening guidelines

Although cervical cancer screening is practiced in many countries, the extent to which it has been organized differs widely.[13] In some countries, women are offered yearly check-ups when they visit their gynecologist, while in others screening is much more structured with an official invitation every 5 years. Especially in settings with non-organized screening, the efficiency and effectiveness of screening could largely be increased by minimizing the extent to which women are over-screened and under-screened.

Reducing over-screening for cervical cancer

In literature, it has been well-acknowledged that over-screening with primary cytology is not beneficial for women. A US modeling study showed that annual cytology screening provides similar life years gained as 3-yearly cytology screening, but requires more than twice the number of colposcopies.[14] Our model outcomes show that annual screening does provide marginal health benefits compared with 3-yearly screening, but comes with roughly 70% more colposcopies (**Chapter 3**). Although the impact on quality of life of certain events

is hard to measure, having a positive cytological test and being referred for colposcopy are certainly associated with adverse psychological effects.[15,16]

Despite the limited or absent benefits, screening at shorter intervals than recommended by international guidelines is still current practice for a lot of women worldwide. In the US, where 3-yearly cytology is recommended, more than 68% of physicians recommend another cytological test in 1 or 2 years,[17] even though more than half of the women are willing to have the 3-year interval.[18] In electronic health record data, 66% of US women aged 30-65 had a shorter screening interval than recommended. Such data can be used to identify women who are screened more often than recommended, and to approach these women and/or their physicians about the ineffectiveness and possible harms of such practice.[19]

In Europe, over-screening rates are also significant, especially in countries with opportunistic screening.[20] Annual cytology is even the official policy in some European countries (Germany, Luxembourg, Austria, and Poland).[20] In those countries, balanced screening guidelines and adherence with those guidelines is needed to improve the harm-benefit ratio of screening, and to limit waste of screening resources. As for over-screened women, switching to primary HPV screening can be considered harmful, cytology should remain the preferred testing method (**Chapter 3**).

HPV-vaccinated women will likely be at much lower risk for cervical cancer, so their potential benefit from intensive screening is even smaller.As the oldest HPV-vaccinated cohorts are approaching the start age of cervical cancer screening, more attention should be drawn to the adverse effects of over-screening. Policy makers, but also physicians and women themselves, should make sure that over-screening is kept to a minimum.

Reducing under-screening for cervical cancer

Meanwhile, a large part of cervical cancer incidence and mortality is found in women who were never screened. In the Netherlands, where 5-year screening coverage is ~80%, more than 50% of cervical cancer cases are found in women who never participated in screening.[21] In countries like the Netherlands, where screening is well-organized and covered by health insurance, organizational barriers are of little importance for (non-)adherence to guidelines. Most non-attendance was found to be related to women's beliefs about cervical screening.[22] Although an informed choice to not participate in screening should be fully respected, the beliefs also included a low perceived likelihood of developing the disease and the belief that the disease is incurable.[22] For those women, awareness campaigns could increase their knowledge and could help them to make an informed decision about screening participation. Although the aversion of the test procedure was not found to be associated with uptake of cervical screening,[22] Gök et al. did show that some never-screened women do participate when offered a self-sampling kit.[23]

In **Chapter 6** of this thesis, we have evaluated the effectiveness of offering self-sampling to non-attendees within the new Dutch cervical cancer screening program, which has started in January 2017.. At the time of that study, policy makers considered sending self-sampling kits to all non-attendees of the screening program. In the program that will be

implemented, however, self-sampling kits will only be sent to women who actively request one. This change in policy reduces the expected increase in participation by 42%,[24] meaning that self-sampling is expected to increase participation with only ~3% of the screening eligible population. This implies that, if self-sampling provides less accurate test results than office-based sampling, it is very important that women do not switch from office-based to self-sampling HPV testing. However, screening will be performed using a polymerase chain reaction (PCR)-based testing method, for which it has been shown that self-sampling and office-based sampling are likely to have similar test accuracy.[25] Nevertheless, switching from office-based to self-sampling should be discouraged as long as a worse test accuracy cannot be ruled out.

Improving adherence with colorectal cancer screening guidelines

For colorectal cancer screening, reducing the percentage of non-attendees is also one of the most important ways to increase the effectiveness of screening. The US Preventive Services Task Force (USPSTF) recommends a variety of possible screening tests,[10] such that everyone can be screened according to their preferences, and the overall uptake is as high as possible. Indeed, providing a range of screening modalities most likely increases uptake with colorectal cancer screening.[26] However, the possibility of multiple screening options was also found to be associated with higher levels of confusion, and individuals who said to be confused by the screening options were less likely to participate in colorectal cancer screening.[27]

Another drawback of offering multiple screening options is that not every test has the same test characteristics, and some tests simply outperform others. Although it is good to stimulate adherence via a range of different tests, one should be careful with offering a new test of which the characteristics are worse than those of other tests. Individuals who would otherwise be screened with a test with a better performance, might switch to the new test, and then overall effectiveness of screening may drop. This might sound hypothetical, but will be true for any individual who switches to a screening strategy with a slightly reduced effectiveness, such as from annual FIT to 3-yearly mtSDNA screening.

New screening tests can only be marketed after approval by the US Food and Drug Administration (FDA). The FDA judges the safety and accuracy of the tests, but does not consider their suitability for mass screening. This suitability is judged by organizations like the USPSTF. If such organizations would not recommend screening tests with worse performance, effectiveness of cancer screening would never decline with introduction of new screening methods. In April 2016, however, the FDA approved the blood test Epi proColon®, a molecular DNA test that detects methylated Septin9 DNA in blood, with an estimated sensitivity of 48% for colorectal cancer.[28] Although the use of such a test would be effective and cost-effective compared to no screening, it does not provide a good alternative to currently available testing methods.[29] Nevertheless, the test is one of the currently available options for colorectal cancer screening in the US, where individuals can choose the screening test with which they want to be screened. The choice of test is not solely based on test characteristics and invasiveness, but can also be influenced by e.g. (aggressive) marketing campaigns.

Possibilities of personalizing screening

The benefit-harm ratio of screening can also be improved by increasing the level of personalization in existing screening programs. To a certain extent, all cancer screening programs are somewhat personalized in that they only target a certain age group in the population. Clearer examples of personalized screening are the intensified breast cancer screening that is offered to women who carry one of the BRCA mutations, and the intensified colorectal cancer screening that is offered to individuals with Lynch Syndrome.

Polygenic testing on the presence of common genetic variants may be used to stratify the average risk population into distinct risk groups. For colorectal cancer, the current discriminatory performance of such testing is insufficient, but this does not mean there is no future for colorectal cancer screening based on polygenic risk. Compared to colorectal cancer, more research has been done to identify single nucleotide polymorphisms (SNPs) associated with breast cancer risk. The AUC value of polygenic testing using 77 SNPs for breast cancer risk is currently ~0.68, and may increase with more GWAS underway, even though the discovery of additional SNPs will yield diminishing returns.[30] The research on SNPs associated with colorectal cancer risk started somewhat later, and might show the same increase in discriminative performance. Our analysis has shown that with an increase in AUC value, the cost-effectiveness of risk-stratified screening based on common genetic variants will rise rapidly (**Chapter 10**).

In addition to including more SNPs, other risk factors for colorectal cancer, such as those related to lifestyle, could be included in risk estimates. A recent study estimated the difference in AUC value between risk-stratification using only SNPs versus using SNPs plus inflammatory bowel disease status and lifestyle risk factors including alcohol consumption, BMI, red meat consumption, smoking, fruit and vegetable intake, physical activity and aspirin usage.[31] Results of this analysis were disappointing; the addition of these risk factors to polygenic risk only led to an increase in the AUC value of 0.01.[31] Another study even found no impact of adding other risk factors (e.g. having a family history) to SNPs.[32] The zero or limited increase could be due to lifestyle factors being similar within families, although literature has shown that family history and lifestyle do independently affect colorectal cancer risk.[33] More research in this field might therefore increase the AUC value of risk-stratification strategies incorporating both polygenic risk, family history, and lifestyle and environmental factors.

Prior to implementation of promising methods to individualize screening, it is important to estimate their effect on the adherence with screening guidelines. The increase in effectiveness of using more individualized screening might easily be offset by a loss in overall adherence with screening. Individualized screening often complicates the process of informed decision making, and this may lead to less intensive screening.[34] On the contrary, three trials have shown that including personalized risk estimates in communication interventions for screening programs enhances informed decision making.[35] There is weak evidence that personalized screening results in higher uptake of screening.[35]

Avoidance of more complicated guidelines could well be a reason for not adopting the suggested colorectal cancer screening based on someone's family history (**Chapter 9**). However, we feel that the increased complexity will be accepted more easily by those with a positive family history. These individuals have seen their family member(s) being diagnosed with the disease, and are often highly motivated to prevent themselves from getting colorectal cancer. For these individuals, too intensive screening may be a larger problem than non-adherence. These people should be made aware that colonoscopies are not just burdensome, but can also result in (fatal) complications. As individuals with a genetic predisposition (which could be indicated by affected family members) generally develop the disease earlier in life, it is important to initiate screening at a younger age, but if no adenomas develop, then the screening interval should be lengthened to minimize unnecessary harm from screening.

FUTURE RESEARCH

Although the work described in this thesis enabled us to answer multiple research questions, more questions keep arising in a dynamic field of research like cancer screening. Some areas for further research are described below.

Possibilities for personalized screening for colorectal cancer

In **Chapter 9** and **10**, we have optimized colorectal cancer screening based on family history and polygenic risk, respectively. Ideally, screening is based on someone's risk of developing the disease within the coming years. For most diseases, this risk is dependent on multiple factors. For colorectal cancer, a family history of the disease and the presence of common genetic variants are just two of those factors. Both are indicators of someone's genetic predisposition to colorectal cancer. Whether someone will develop colorectal cancer, also depends on his/her lifestyle, environment, the unexplainable 'bad luck' factor, and someone's screening history. Future research should aim at identifying full risk profiles of individuals, including as much relevant information as possible.

The expected benefit of cancer screening not only depends on the probability that a cancer diagnosis or death is averted, but also on the expected gain in (healthy) life years. The expected harm-benefit ratio of screening is therefore dependent on someone's (healthy) life expectancy. There is general acceptance that for individuals with limited life expectancy, the net benefit of screening is either small or negative.[36] Current co-morbidity status and risk for other diseases are therefore important to incorporate in personalized screening.

Earlier simulation modeling showed that the optimal end age of colonoscopy screening varies widely with screening history, background colorectal cancer risk and co-morbidity status.[37] It would be interesting to expand this work, by optimizing the optimal start age and interval of screening, and by including also other screening modalities than colonoscopy. By comparing modeling outcomes of such personalized screening with uniform screening,

we would get a more precise estimate of the potential usefulness of personalized screening for colorectal cancer.

Possibilities for personalized screening for cervical cancer

For cervical cancer, one may consider offering risk-stratified screening based on a woman's risk to acquire a high-risk HPV infection. Although this is unsuitable for unvaccinated cohorts, in which sexual behavior is the main factor influencing the risk of acquiring HPV, it might be fairly straightforward for cohorts that have been offered vaccination. Due to large expected differences in cervical cancer risk, tailored screening based on vaccination status is likely cost-effective (**Chapter 8**). Although cigarette smoking[38,39], long-term oral contraceptive use[39,40], and multiparity[39] were found to be associated with progression to CIN 3, their correlation with progression to cervical cancer is uncertain. More research is needed to identify factors that are associated with an increased (or decreased) cervical cancer risk.

Just as with colorectal cancer screening, it is also possible to personalize screening based on co-morbidity status. Although cervical cancer screening is offered at younger ages than colorectal cancer screening, comorbid conditions may still influence the expected (cost-) effectiveness of screening, especially at ages 60-70.

Issues with personalization of screening

As said, in future cohorts, the most obvious direction for risk-stratified cervical cancer screening would be screening based on HPV-vaccination status. Although this has shown to be cost-effective, it is uncertain whether vaccinated women will agree with being offered less intensive screening, just because they adhered to vaccination guidelines. Such a reduction in screening frequency will be difficult to implement for any type of screening, and for any type of reason. To increase the acceptability of the population, the reasoning behind the risk-stratification should be explained very clearly.

Although in theory, risk-stratified screening is always beneficial because it involves a more efficient way of allocating screening resources, this does not have to hold in practice. If risk-stratified screening only leads to individuals at increased risk being screened more often and not to those at decreased risk being screened less often, it is uncertain whether projected benefits will be achieved. Therefore, and because cancer risk is overrepresented in individuals who do not attend screening, attempts to increase adherence with current screening guidelines by e.g. offering self-sampling to non-attendees, are likely more effective than currently available possibilities for risk-stratified screening.

Attempts to increase adherence

Cervical and colorectal cancer screening are effective, but could be enhanced by an increased adherence with screening guidelines. In 2013, 81% of screening eligible US women reported to have had a cytological test in the past three years.[48] At 59%, adherence with either endoscopy or FOBT for colorectal cancer screening was much lower.[48] For both

cervical and colorectal cancer screening, uptake increases with level of education, varies by ethnicity and is significantly higher in individuals with health insurance.[48]

Lower participation with colorectal cancer screening than with cervical cancer screening can partly be explained by differences in target population. Men are less likely to adhere with colorectal cancer screening guidelines than women.[49] Women tend to be more knowledgeable about screening, and are therefore better at making informed decisions about screening.[49] If women delay or refuse screening, it is mostly because of aversion to the testing procedure or perceived distress from screening. Men tend to procrastinate making an informed decision about screening, and often deny its possible importance.[49] Adherence with screening could therefore be increased by supporting men in making definitive decisions about screening, and by taking away women's anxiety around colorectal cancer screening.[49]

Although colorectal cancer screening uptake is ~5% higher in women than in men, this finding cannot explain the 22 percent point difference in uptake between cervical and colorectal cancer screening. A study among UK women who were eligible for both types of screening revealed that dislike of the test seems to be a larger barrier to colorectal cancer screening than to breast and cervical cancer screening.[50] However, even with the wide range of screening tests available in the US, uptake with colorectal cancer screening is relatively low. More research is needed to identify perceived barriers to this type of screening.

For cervical cancer screening, self-sampling tests require less time than office-based tests and eliminate the involvement of a clinician, and are therefore expected to increase screening participation.[23] However, also self-sampling requires taking a cervical smear, which can be considered as an invasive procedure. Recent studies have shown that HPV testing in urine also has good accuracy for the detection of HPV and cervical lesions.[51,52] More research on such testing methods is needed to determine whether it also provides sufficient protection against invasive cancer.

Personalized treatment

While cancer treatment is getting more personalized, treatment of pre-invasive lesions is often still regarded as general practice. Guidelines generally recommend that CIN 1 should not be treated directly. Given the long preclinical phase of progressive disease, the lesion can still be treated if it does not regress within e.g. 1-3 years. Some local guidelines also acknowledge that direct treatment of CIN 2 may not be the best option for all women. WHO guidelines do recommend treatment of CIN 2+ for all women, regardless of age and potential childwish.[41] However, with a regression rate of ~70% at 3 years after diagnosis, regression of CIN 2 is also common.[42] This means that active surveillance may also be appropriate for women with CIN 2, especially for those with a (future) childwish, given that treatment increases their risk of having subsequent preterm labor. However, the probability of preterm birth significantly increases with the invasiveness of the treatment.[5] This means that if the size of the lesion increases instead of reduces, active surveillance instead of direct treatment may result in the need for a larger cone depth, which could result in an increased

potential harm from treatment. It would be interesting to model the effect of delaying the treatment of <CIN 2 and <CIN 3 in terms of preterm births and averted cervical cancer diagnoses and deaths.

As evidence for regression of colorectal adenomas is limited, guidelines state that any detected adenoma should be removed. However, the average preclinical duration from a small adenoma to clinical cancer is >10 years,[43] and treatment is associated with risk of serious complications, i.e. gastrointestinal and cardiovascular events, which increases exponentially with age.[44] Therefore, the expected net benefit of removing a small adenoma in individuals with limited life expectancy is probably small and might even be negative. Future modeling studies may estimate the potential benefit of more restrained treatment policies for small adenomas.

For both cervical lesions and colorectal adenomas, one would ideally be able to identify which precursor lesions will develop into clinical cancer, and within what timeframe. If such identification would be possible, then only those CIN and adenomas would require treatment, and overtreatment could largely be reduced. For cervical cancer, multiple studies have shown that p16 and L1 may be promising markers of progression of low-grade cytological abnormalities.[45,46] For colorectal cancer, several candidate biomarkers for adenoma-to-carcinoma progression have been identified.[47] More research is needed to improve prediction of which lesions are likely to progress and should therefore be treated.

CONCLUSIONS

From the results of the studies that are presented in this thesis, the following conclusions can be drawn:

- Many published cost-effectiveness analyses on cervical cancer screening are likely to have omitted relevant comparator strategies, thereby possibly identifying an inefficient screening strategy as optimal.
- For over-screened women, switching from primary cytology to primary HPV screening slightly reduces cervical cancer risk but comes with a large increase in the number of (false-)positive tests, and can therefore be considered as harmful.
- Cervical cancer screening from age 21 instead of age 30 only slightly reduces cervical cancer risk, while it substantially increases the risk of preterm birth.
- Data from the Dutch cervical cancer screening program suggest that SurePath detects more progressive lesions than conventional cytology, while ThinPrep might detect fewer.
- Offering a self-sample test to non-attendees in a cervical cancer screening program is likely a cost-effective intervention, but its impact does depend on parameters like the relative sensitivity of the self-sample test (as compared to the regular test), the extent to which it generates extra attendance, and the extent to which attendees of regular screening switch to using the self-sample test.

- Although screening with the multitarget stool DNA test is an effective way to reduce colorectal cancer incidence and mortality, the current price of the test makes it a cost-ineffective alternative to other stool-based tests.
- When herd immunity reaches beyond 50%, offering unvaccinated women cervical cancer screening based on the pre-vaccination risk level is no longer cost-effective, and reducing the intensity of uniform screening may be considered.
- It is cost-effective to offer individuals with a family history of colorectal cancer intensive colonoscopy screening from a young age onwards. However, if individuals do not develop any adenomas, it is unlikely that they have a genetic predisposition to colorectal cancer, so continuing intensive screening would provide little or no additional benefit.
- At the discriminatory performance and cost of current polygenic tests, colonoscopy screening based on polygenic risk is not cost-effective compared to uniform screening. This may change with a reduction in the price per polygenic test and with the discovery of more common genetic variants that are associated with colorectal cancer risk.

RECOMMENDATIONS

The conclusions derived in this thesis support the following recommendations:

- In a cancer screening cost-effectiveness analysis, one should include all relevant comparator strategies to obtain reliable cost-effectiveness estimates.
- In screening situations that are characterized by a substantial amount of over-screening, primary cytology screening should not be replaced by primary HPV screening. Reducing the amount of over-screening should be prioritized.
- Cervical cancer screening below age 30 should only be offered after careful evaluation of a woman's expected benefits and harms, taking into account her (future) childwish.
- Since 2017, primary LBC screening is longer performed in the Netherlands, but the fact that the use of SurePath resulted in different screening outcomes than the use of ThinPrep should raise international awareness that the assumed similarity in test characteristics may not be true.
- Although offering an alternative test could increase effectiveness of screening through increasing the uptake among individuals willing to undergo screening, one should be careful with offering alternative tests with inferior characteristics compared to already offered tests because of the potential substitution effect.
- If cervical cancer screening cannot be tailored to HPV-vaccination status, one should wait with adjusting the screening frequency to the risk level in vaccinated women until unvaccinated women benefit from a herd immunity level of at least 50%, i.e. until the reduction in HPV prevalence in unvaccinated women is at least half of that in vaccinated women.
- As screening intensity preferably depends on the disease risk in the target population, herd immunity should be closely monitored in partly vaccinated cohorts.

- Individuals with a constant level of family history of colorectal cancer should consider lengthening their screening interval after several subsequent negative colonoscopies, as this suggests that they are unlikely to have a genetic predisposition for colorectal cancer.
- In addition to discovering more common genetic variants that are associated with colorectal cancer risk, research should aim at identifying comprehensive colorectal cancer risk profiles, including both genetic and environmental risk factors.
- Given high regression rates of CIN 1 and CIN 2, and the potential adverse effects of CIN treatment, future research is needed to estimate the impact of delaying treatment in women with CIN 1 and CIN 2.
- Much of the harm related to cervical and colorectal cancer screening comes from the detection and treatment of non-progressive lesions. Literature suggests the existence of biomarkers that are able to distinguish between progressive and non-progressive lesions. As this may largely reduce the amount of overdiagnosis and overtreatment, more research is needed to identify those biomarkers.

REFERENCES

1. Ronco G, Dillner J, Elfstrm KM, et al. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. Lancet 2014;383:524-32.
2. Australian Government: Department of Public Health. National Cervical Screening Program: Future changes to cervical screening.
3. National Institute for Public Health and the Environment. [Screening program will change] Bevolking-sonderzoek gaat veranderen.
4. Kyrgiou M, Tsoumpou I, Vrekoussis T, et al. The up-to-date evidence on colposcopy practice and treatment of cervical intraepithelial neoplasia: the Cochrane colposcopy & cervical cytopathology collaborative group (C5 group) approach. Cancer Treat Rev 2006;32:516-23.
5. Danhof NA, Kamphuis EI, Limpens J, van Lonkhuijzen LRCW, Pajkrt E, Mol BWJ. The risk of preterm birth of treated versus untreated cervical intraepithelial neoplasia (CIN): a systematic review and meta-analysis. Eur J Obstet Gynecol Reprod Biol. 2015 May;188:24-33.
6. Hodek JM, von der Schulenburg JM, Mittendorf T. Measuring economic consequences of preterm birth - Methodological recommendations for the evaluation of personal burden on children and their caregivers. Health Econ Rev 2011;1:6.
7. Arbyn M, Bergeron C, Klinkhamer P, et al. Liquid compared with conventional cervical cytology: a systematic review and meta-analysis. Obstet Gynecol 2008;111:167-77.
8. Whitlock EP, Vesco KK, Eder M, et al. Liquid-based cytology and human papillomavirus testing to screen for cervical cancer: a systematic review for the U.S. Preventive Services Task Force. Ann Intern Med 2011;155:687-97, W214-5.
9. Rozemeijer K, Penning C, Siebers AG, et al. Comparing SurePath, ThinPrep, and conventional cytology as primary test method: SurePath is associated with increased CIN II(+) detection rates. Cancer Causes Control 2016;27:15-25.
10. U. S. Preventive Services Task Force, Bibbins-Domingo K, Grossman DC, et al. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. JAMA 2016;315:2564-75.
11. Ramsey SD, Wilschut J, Boer R, et al. A decision-analytic evaluation of the cost-effectiveness of family history-based colorectal cancer screening programs. Am J Gastroenterol 2010;105:1861-9.
12. Wilschut JA, Steyerberg EW, van Leerdam ME, et al. How much colonoscopy screening should be recommended to individuals with various degrees of family history of colorectal cancer? Cancer 2011;117:4166-74.
13. Elfström KM, Arnheim-Dahlström L, von Karsa L, et al. Cervical cancer screening in Europe: Quality assurance and organisation of programmes. Eur J Cancer 2015;51:950-68.
14. Kulasingam SL, Havrilesky L, Ghebre R, et al. Screening for Cervical Cancer: A Decision Analysis for the U.S. Preventive Services Task Force. 2011.
15. O'Connor M, Gallagher P, Waller J, et al. Adverse psychological outcomes following colposcopy and related procedures: a systematic review. BJOG 2016;123:24-38.
16. Lerman C, Miller SM, Scarborough R, et al. Adverse psychologic consequences of positive cytologic cervical screening. Am J Obstet Gynecol 1991;165:658-62.
17. Saraiya M, Berkowitz Z, Yabroff KR, et al. Cervical cancer screening with both human papillomavirus and Papanicolaou testing vs Papanicolaou testing alone: what screening intervals are physicians recommending? Arch Intern Med 2010;170:977-85.
18. Meissner HI, Tiro JA, Yabroff KR, et al. Too much of a good thing? Physician practices and patient willingness for less frequent pap test screening intervals. Med Care 2010;48:249-59.

19. Mathias JS, Gossett D, Baker DW. Use of electronic health record data to evaluate overuse of cervical cancer screening. Journal of the American Medical Informatics Association 2012;19:E96-E101.
20. Arbyn M, Rebolj M, De Kok IM, et al. The challenges of organising cervical screening programmes in the 15 old member states of the European Union. Eur J Cancer 2009;45:2671-8.
21. Bos AB, Rebolj M, Habbema JD, et al. Nonattendance is still the main limitation for the effectiveness of screening for cervical cancer in the Netherlands. Int J Cancer 2006;119:2372-5.
22. Tacken MA, Braspenning JC, Hermens RP, et al. Uptake of cervical cancer screening in The Netherlands is mainly influenced by women's beliefs about the screening and by the inviting organization. Eur J Public Health 2007;17:178-85.
23. Gök M, Heideman DA, van Kemenade FJ, et al. Offering self-sampling for human papillomavirus testing to non-attendees of the cervical screening programme: Characteristics of the responders. Eur J Cancer 2012;48:1799-808.
24. Verdoodt F, Jentschke M, Hillemanns P, et al. Reaching women who do not participate in the regular cervical cancer screening programme by offering self-sampling kits: a systematic review and meta-analysis of randomised trials. Eur J Cancer 2015;51:2375-85.
25. Arbyn M, Snijders PJ, Meijer CJ, et al. Which high-risk HPV assays fulfil criteria for use in primary cervical cancer screening? Clin Microbiol Infect 2015;21:817-26.
26. Marshall DA, Johnson FR, Phillips KA, et al. Measuring patient preferences for colorectal cancer screening using a choice-format survey. Value Health 2007;10:415-30.
27. Jones RM, Vernon SW, Woolf SH. Is discussion of colorectal cancer screening options associated with heightened patient confusion? Cancer Epidemiol Biomarkers Prev 2010;19:2821-5.
28. Church TR, Wandell M, Lofton-Day C, et al. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. Gut 2014;63:317-25.
29. Ladabaum U, Allen J, Wandell M, et al. Colorectal cancer screening with blood-based biomarkers: cost-effectiveness of methylated septin 9 DNA versus current strategies. Cancer Epidemiol Biomarkers Prev 2013;22:1567-76.
30. Wu Y, Abbey CK, Liu J, et al. Discriminatory power of common genetic variants in personalized breast cancer diagnosis. Proc SPIE Int Soc Opt Eng 2016;9787.
31. Frampton M, Houlston RS. Modeling the prevention of colorectal cancer from the combined impact of host and behavioral risk factors. Genet Med 2016.
32. Hsu L, Jeon J, Brenner H, et al. A model to determine colorectal cancer risk using common genetic susceptibility loci. Gastroenterology 2015;148:1330-9 e14.
33. Huang XE, Hirose K, Wakai K, et al. Comparison of lifestyle risk factors by family history for gastric, breast, lung and colorectal cancer. Asian Pac J Cancer Prev 2004;5:419-27.
34. Lieberman D. Progress and challenges in colorectal cancer screening and surveillance. Gastroenterology 2010;138:2115-26.
35. Edwards AG, Naik G, Ahmed H, et al. Personalised risk communication for informed decision making about taking screening tests. Cochrane Database Syst Rev 2013:CD001865.
36. Royce TJ, Hendrix LH, Stokes WA, et al. Cancer screening rates in individuals with different life expectancies. JAMA Intern Med 2014;174:1558-65.
37. van Hees F, Habbema JD, Meester RG, et al. Should colorectal cancer screening be considered in elderly persons without previous screening? A cost-effectiveness analysis. Ann Intern Med 2014;160:750-9.
38. Roura E, Castellsague X, Pawlita M, et al. Smoking as a major risk factor for cervical cancer and pre-cancer: results from the EPIC cohort. Int J Cancer 2014;135:453-66.
39. Luhn P, Walker J, Schiffman M, et al. The role of co-factors in the progression from human papillomavirus infection to cervical cancer. Gynecol Oncol 2013;128:265-70.

40. Gierisch JM, Coeytaux RR, Urrutia RP, et al. Oral contraceptive use and risk of breast, cervical, colorectal, and endometrial cancers: a systematic review. Cancer Epidemiol Biomarkers Prev 2013;22:1931-43.
41. World Health Organization. WHO guidelines for treatment of cervical intraepithelial neoplasia 2–3 and adenocarcinoma in situ: cryotherapy, large loop excision of the transformation zone, and cold knife conization. 2014. http://apps.who.int/iris/bitstream/10665/94830/1/9789241548694_eng.pdf Accessed February 27, 2017.
42. Moscicki AB, Ma Y, Wibbelsman C, et al. Rate of and risks for regression of cervical intraepithelial neoplasia 2 in adolescents and young women. Obstet Gynecol 2010;116:1373-80.
43. Morson B. President's address. The polyp-cancer sequence in the large bowel. Proc R Soc Med 1974;67:451-7.
44. van Hees F, Zauber AG, Klabunde CN, et al. The Appropriateness of More Intensive Colonoscopy Screening Than Recommended in Medicare Beneficiaries - A Modeling Study. JAMA Intern Med 2014;174:1568-1576.
45. Ungureanu C, Socolov D, Anton G, et al. Immunocytochemical expression of p16INK4a and HPV L1 capsid proteins as predictive markers of the cervical lesions progression risk. Rom J Morphol Embryol 2010;51:497-503.
46. Negri G, Bellisano G, Zannoni GF, et al. p16 ink4a and HPV L1 immunohistochemistry is helpful for estimating the behavior of low-grade dysplastic lesions of the cervix uteri. Am J Surg Pathol 2008;32:1715-20.
47. de Wit M, Jimenez CR, Carvalho B, et al. Cell surface proteomics identifies glucose transporter type 1 and prion protein as candidate biomarkers for colorectal adenoma-to-carcinoma progression. Gut 2012;61:855-64.
48. Smith RA, Andrews K, Brooks D, et al. Cancer screening in the United States, 2016: A review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin 2016;66:96-114.
49. Sinicrope PS, Goode EL, Limburg PJ, et al. A population-based study of prevalence and adherence trends in average risk colorectal cancer screening, 1997 to 2008. Cancer Epidemiol Biomarkers Prev 2012;21:347-50.
50. Lo SH, Waller J, Wardle J, et al. Comparing barriers to colorectal cancer screening with barriers to breast and cervical screening: a population-based survey of screening-age women in Great Britain. J Med Screen 2013;20:73-9.
51. Pathak N, Dodds J, Zamora J, et al. Accuracy of urinary human papillomavirus testing for presence of cervical HPV: systematic review and meta-analysis. BMJ 2014;349:g5264.
52. Piyathilake CJ, Badiga S, Chambers MM, et al. Accuracy of urinary human papillomavirus testing for the presence of cervical human papillomaviruses and higher grades of cervical intraepithelial neoplasia. Cancer 2016;122:2836-44.

Model appendix

INTRODUCTION

MISCAN is a stochastic, semi-Markov microsimulation model programmed in Delphi (Borland Software Corporation, Scotts Valley, California, United States). It can be used to explain and predict trends in cancer incidence and mortality and to quantify the effects and costs of primary prevention, screening, and surveillance.

The term 'microsimulation' implies that individuals are moved through the model one at a time, rather than as proportions of a cohort. This allows future state transitions to depend on past transitions, giving the model a 'memory'. Furthermore, unlike most traditional Markov models, MISCAN does not use yearly or monthly transition probabilities; instead it generates durations in states, thereby increasing model flexibility and computational performance. The term 'stochastic' implies that the model simulates sequences of events by drawing from distributions of probabilities/durations, rather than using fixed values. Hence, the results of the model are subject to random variation.

In this thesis, two different types of the MISCAN model have been used, i.e. MISCAN-Cervix for the evaluation of cervical cancer screening, and MISCAN-Colon for the evaluation of colorectal cancer (CRC) screening. Both MISCAN-Cervix and MISCAN-Colon consist of 3 modules: a demography module, a natural history module, and a screening module (**Figure 1**), of which the assumptions are described in the next two sections.

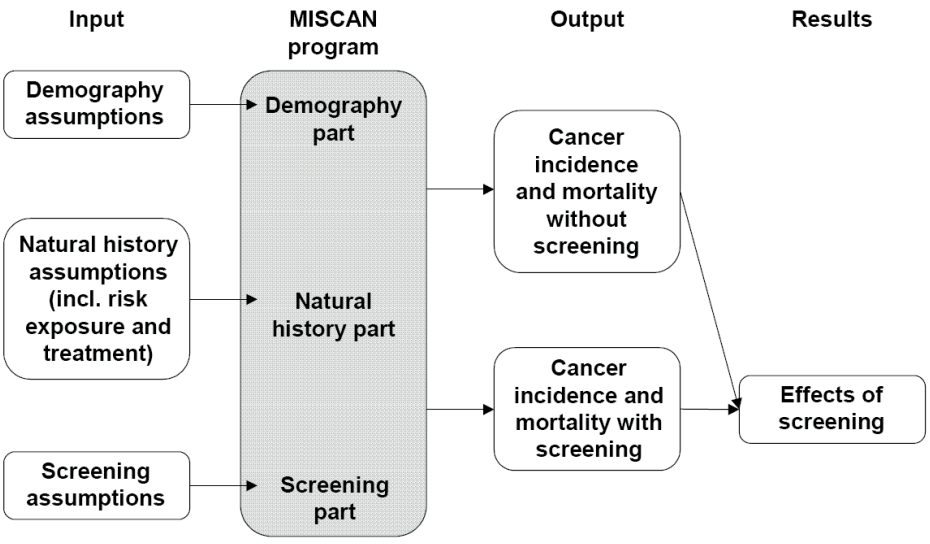


Figure 1. Structure of the MISCAN model.

MODULES IN MISCAN-CERVIX

Demography module

Using birth tables and life tables representative for the female population under consideration, MISCAN-Cervix draws a date of birth and a date of death from other causes for each simulated individual. In this thesis, we either simulated a cohort of a specific age (**Chapters 3**, **4**, and **8**) or a population with an age distribution representative for the Dutch female population (**Chapter 6**). Life tables were retrieved from Statistics Netherlands.[1] In MISCAN-Cervix the maximum age a woman can achieve is exactly 100 years.

Women who have had a total hysterectomy (i.e. full removal of both the uterus and the cervix) can no longer develop cervical cancer. In MISCAN-Cervix, cohort- and age-specific probabilities determine the likelihood of getting such a hysterectomy due to causes other than cervical cancer. In this thesis, the probability of getting a hysterectomy was based on data from Statistics Netherlands and the Information Centre for Health Care.[1,2]

Natural history module

During her lifetime, each woman has an age-specific risk of acquiring high-risk HPV infections (i.e. an infection caused by an HPV type that can cause cancer and that can be detected by the HPV test) and CIN lesions without a (detectable) high-risk HPV infection. Most HPV infections clear or regress naturally, some HPV infections can progress to CIN 1, CIN 2, CIN 3, cervical cancer, and death from cervical cancer.

The age-specific incidence of HPV infections that progress to cervical cancer was calibrated to the age-specific incidence of cervical cancer, which was obtained from the Netherlands Cancer Registry (**Figure 2**).[3] The age-specific incidence of pre-invasive lesions that do not progress to cervical cancer was calibrated so that the simulated detection rates of CIN lesions fit the observed detection rates in the Netherlands (**Figure 3**). The observed detection rates were obtained from the nationwide network and registry of histo- and cytopathology in the Netherlands (PALGA) for the period 2000-2007.[4] The incidence of high-risk HPV infections that do not progress to CIN was calibrated so that the simulated prevalence of all high-risk HPV infections fits the observed high-risk HPV prevalence (**Figure 4**).[5,6]

In MISCAN-Cervix, 6 disease pathways are distinguished. Each instance of these disease pathways represents an HPV infection or a 'lesion' (i.e. CIN of a certain grade or a stage of cervical cancer). Each disease pathway starts as either an HPV infection or as an HPV negative CIN 1 lesion. The natural history (i.e. in the situation without screening) of these 6 disease pathways is shown in **Figure 5** and can be described as follows:

A) HPV infections that clear naturally without ever leading to CIN
B) HPV infections that progress to CIN 1 and then regress
C) HPV infections that progress to CIN 1 and CIN 2 and then regress
D) HPV infections that progress to CIN 1, CIN 2, and CIN 3 and then regress

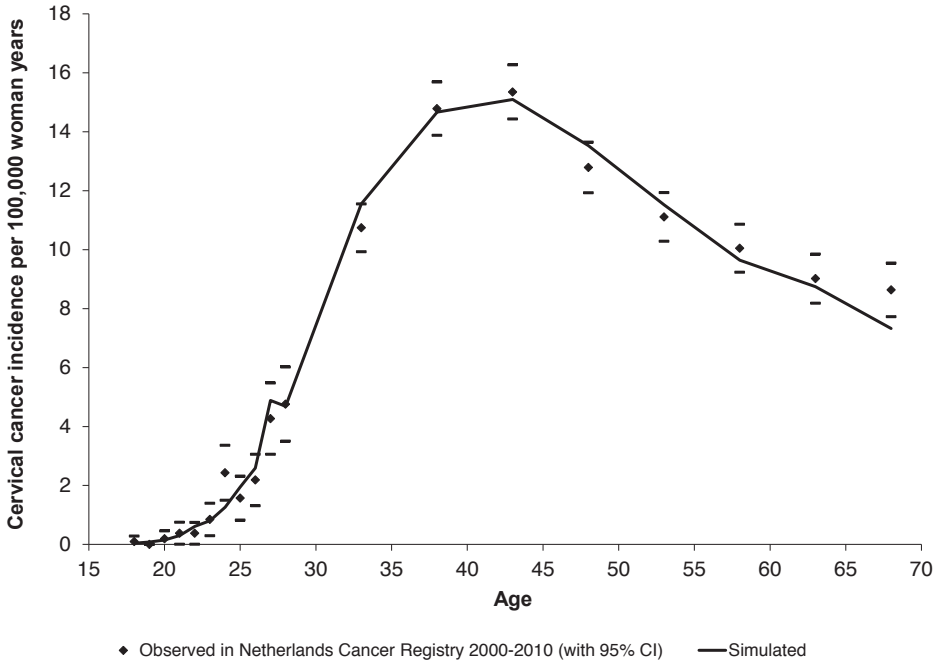


Figure 2. Cervical cancer incidence as observed in the Netherlands Cancer Registry in 2000-2010 versus simulated by MISCAN-Cervix (cases per 100,000 woman years).

E) HPV negative CIN 1 lesions that regress naturally or become HPV negative CIN 2 and then regress naturally
F) HPV infections that progress to CIN 1, CIN 2, CIN 3, preclinical cervical cancer stage IA (micro-invasive), and preclinical cervical cancer stage IB. Preclinical cervical cancer stage IB can either become clinically detected cervical cancer stage IB or progress to preclinical cervical cancer stage II+ and then to clinical cervical cancer stage II+. Clinically detected cervical cancer can progress to death from cervical cancer or remain in that state forever (if the woman is cured from cervical cancer).

A woman can acquire multiple lesions and HPV infections during her lifetime, and multiple lesions and HPV infections may be present at the same time. In each simulated life history (i.e. between ages 0 and 100), the number of lesions of each type follows a Poisson distribution. The annual probability of acquiring an HPV infection or CIN lesion is age-dependent. The transitions and sojourn times of the HPV infections or lesions are simulated based on a continuous-time semi-Markov process. The sojourn times of most states in the model have either an exponential or a Weibull probability distribution.

In the model, women who do not have cervical cancer have an age-specific probability of getting a hysterectomy for reasons other than cervical cancer. A hysterectomy is assumed to remove all prevalent HPV infections and CIN lesions. After a hysterectomy, women are

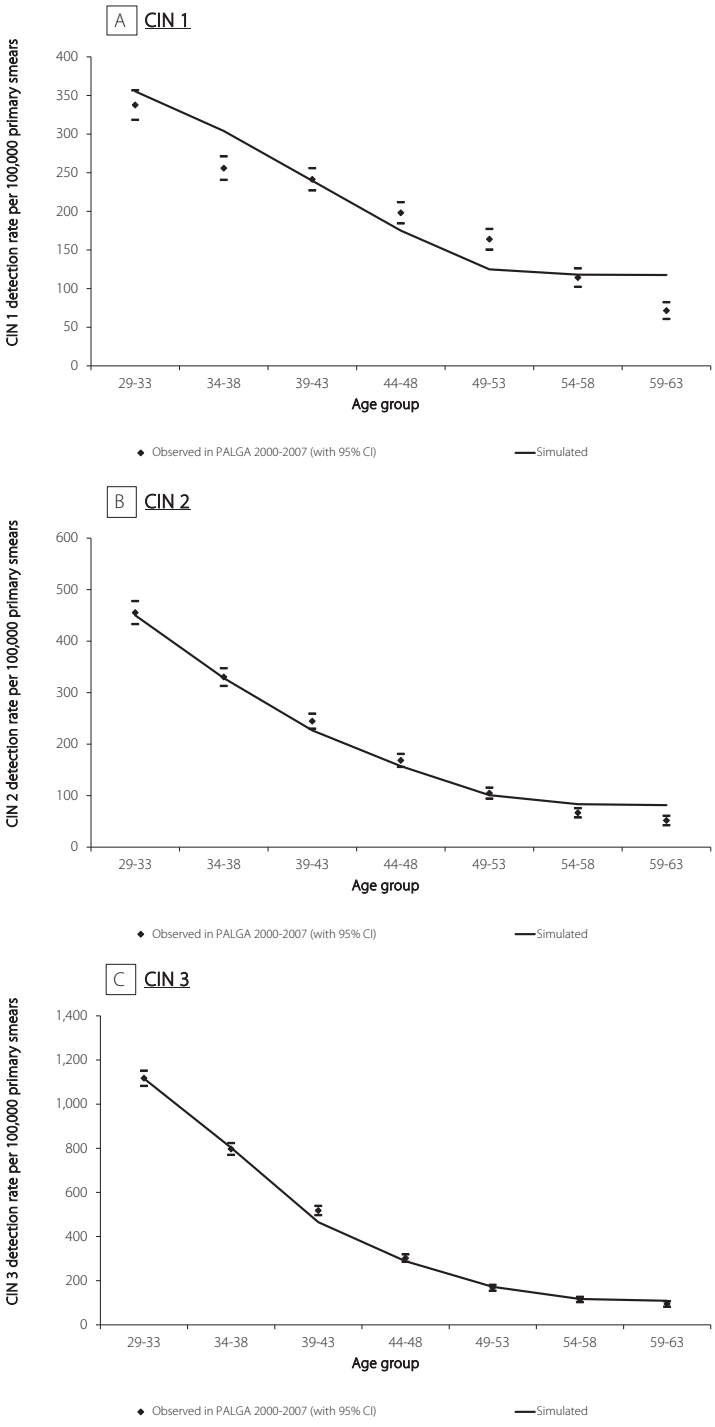


Figure 3. CIN detection rates as observed in PALGA in 2000-2007 versus simulated by MISCAN-Cervix (CIN 1 (A), CIN 2 (B), CIN 3 (C); lesions detected per 100,000 primary smears).
CIN = cervical intraepithelial neoplasia.

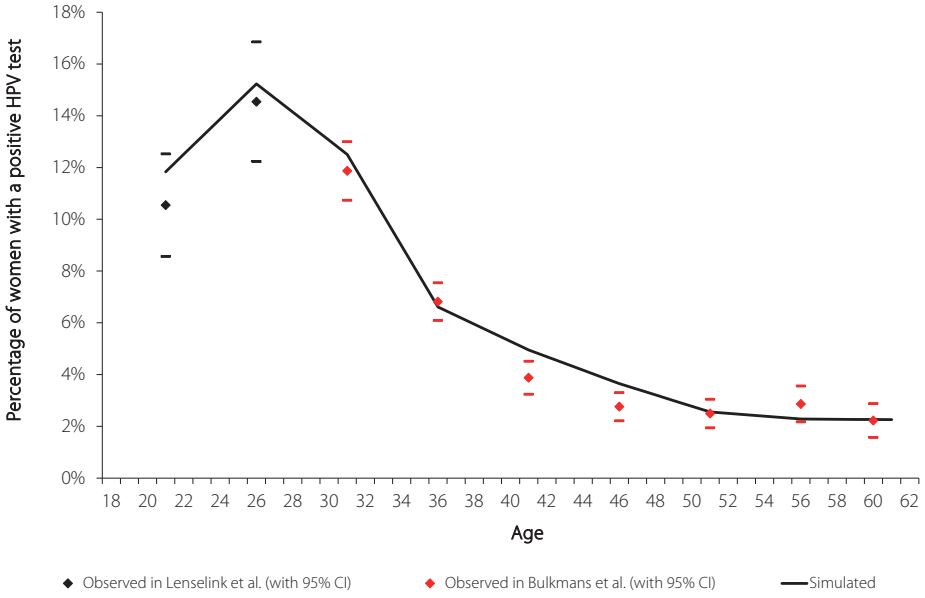


Figure 4. Percentage of women with a positive HPV test as observed in Bulkmans et al. and Lenselink et al. versus simulated by MISCAN-Cervix.
HPV = human papillomavirus.

no longer at risk for HPV infections and CIN lesions, and are therefore no longer invited for screening.

The assumptions for the probability and the duration of survival after a clinically detected (i.e. detected because of symptoms) cervical cancer are based on data from the NCR for the period 1989-2009. As these data include both adenocarcinoma and squamous cell carcinoma, the survival we estimated is a weighted average of these two types of cervical cancer. We assumed that all cervical cancer mortality occurs in the first 10 years after diagnosis. The assumed probability of long-term survival depends on age and stage (IB or II+); in the model, cervical cancer stage IA cannot be clinically detected.

Screening module

Screening will alter some of the simulated life histories: Some cancers will be prevented by the detection and removal of CIN lesions; other cancers will be detected in an earlier stage with a more favorable survival. The simulated screening protocol determines in what situations women are referred for colposcopy. If a woman is referred to colposcopy, all prevalent CIN lesions are assumed to be diagnosed and successfully removed. HPV infections without CIN are not treated. For screen-detected cervical cancer, a stage-specific improvement (compared to the situation without screening) in the probability of cure is assumed.

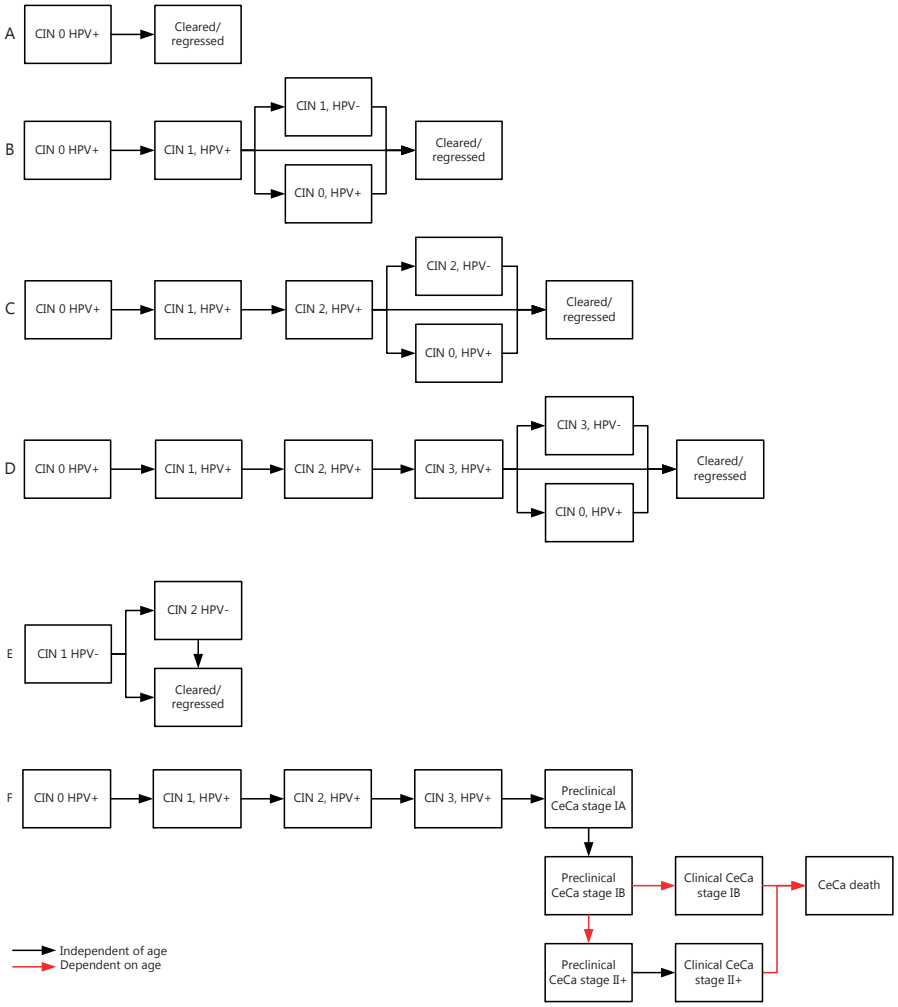


Figure 5. Schematic representation of the natural history module of MISCAN-Cervix, with disease pathways A through F.

HPV = human papillomavirus; CIN = cervical intraepithelial neoplasia; CeCa = cervical cancer.

Note: There are six disease pathways (types A through F) in MISCAN-Cervix. All lesions start as either an HPV infection without CIN (disease pathways A, B, C, D, and F) or as a CIN 1 lesion without HPV infection (disease pathway E). Cleared/regressed denotes the absence of CIN and HPV infection; CIN 0 denotes the absence of CIN and cervical cancer. All cervical cancer states are HPV positive. The arrows between the states show which types of transitions can occur. In every state before death, a transition to "death from other causes" can occur, and in every state before cancer, a transition to "hysterectomy" can occur (states and connecting arrows not shown); in these cases, the transition applies to all HPV infections and CIN lesions of that person simultaneously.

In the model, detection of cervical cancer by screening prevents death from cervical cancer in some but not all cases. However, if death from cervical cancer is not prevented, the time of death from cervical cancer is not changed by screening.

For screen-detected invasive cancers, survival was modeled as a reduction in the risk of dying compared with that risk in the situation without screening, when the cancer would have become clinical. This improvement of prognosis (89.4%, 50% and 20% for stage IA, IB and II+ respectively) was calibrated to reproduce recently observed stage-specific survival given observed screening (Netherlands Cancer Registry).[3]

MODULES IN MISCAN-COLON

Demography module

Using birth tables and life tables representative for the population under consideration, MISCAN-Colon draws a date of birth and a date of death from other causes for each simulated individual. In this thesis, we simulated single birth cohorts to evaluate screening decisions at certain ages. Life tables were retrieved from the Centers for Disease Control and Prevention.[7] In MISCAN-Colon the maximum age an individual can achieve is exactly 100 years.

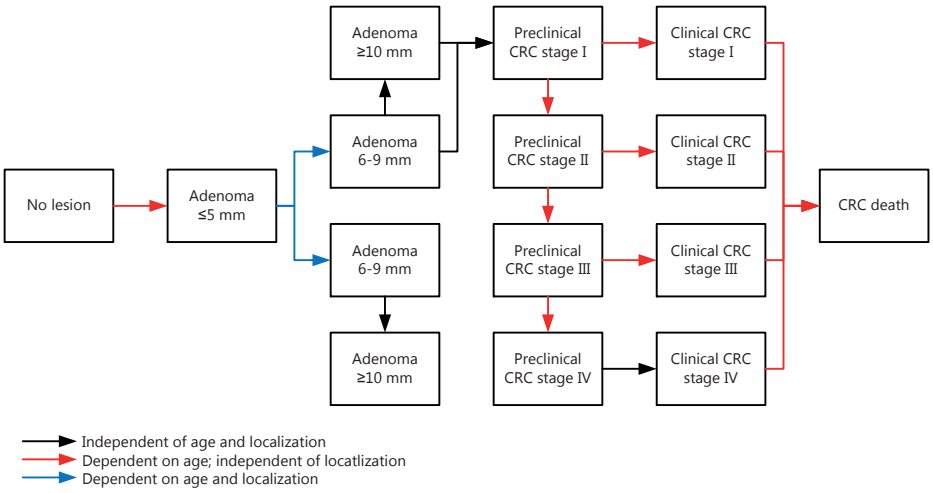


Figure 6. Schematic representation of the natural history module of MISCAN-Colon.
CRC = colorectal cancer.
The arrows between the states show which types of transitions can occur. In every state before death, a transition to "death from other causes" can occur (state and connecting arrows are not shown).

Natural history module

As each simulated person ages, one or more adenomas may develop (**Figure 6**). These adenomas can be either progressive or non-progressive. Both progressive and non-progressive adenomas can grow in size from small (≤5 mm), to medium (6-9 mm), to large (≥10 mm); however, only progressive adenomas can develop into preclinical cancer. A preclinical cancer may progress through stages I to IV; however, during each stage CRC may be diagnosed because of symptoms. After clinical diagnosis, CRC survival is simulated using age-, stage-, and localization-specific survival estimates for clinically diagnosed CRC obtained from a study by Rutter and colleagues.[8] For individuals with synchronous CRCs at time of diagnosis, the survival of the most advanced cancer is used. The date of death for individuals with CRC is set to the earliest simulated death (either due to CRC or due to another cause (see: 'Demography module')).

An individual's risk of developing adenomas depends on the individual's age and a personal risk index. As a result of the latter most individuals develop no adenomas, whilst some develop many. We assumed that the distribution of adenomas over the colon and rectum equals the distribution of cancers as observed in SEER before the introduction of screening.[9] The age-specific onset of adenomas and the dispersion of the personal risk index were calibrated to data on the prevalence and multiplicity distribution of adenomas as observed in autopsy studies (**Figure 7**).[10-19] The age-specific probability of adenoma-progressivity and the age- and localization-specific transition probabilities between preclinical cancer stages

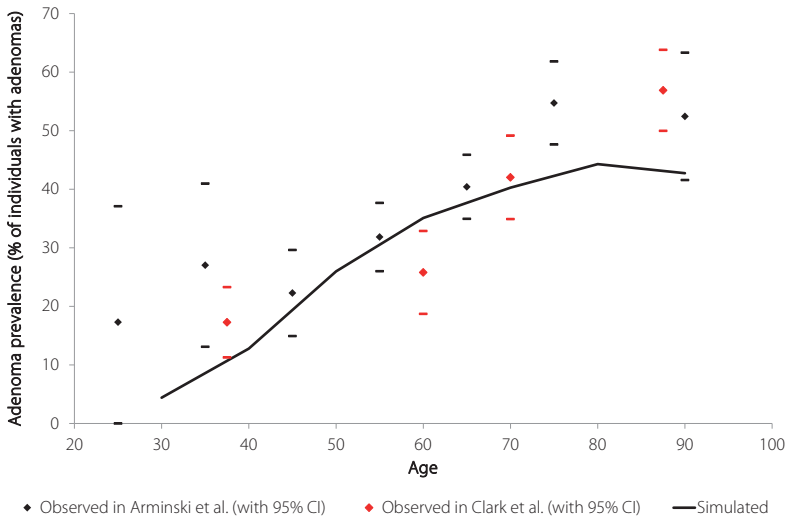


Figure 7. Adenoma prevalence observed in selected autopsy studies versus simulated by MISCAN-Colon (% of individuals with adenomas).*

* Observed results are only shown for the two largest studies (Arminski et al.[10] and Clark et al.[13]) on which the model has been calibrated. MISCAN-Colon has additionally been calibrated to 8 other autopsy studies.[9,11,12,14-18]

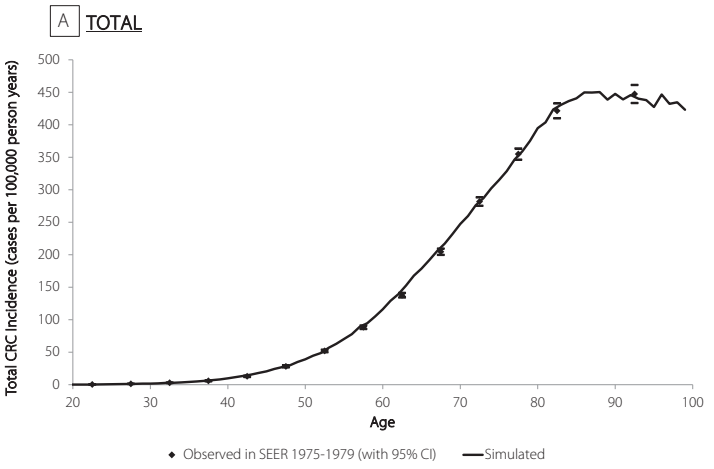
and between preclinical and clinical cancer stages were simultaneously calibrated to SEER data on the age-, stage-, and localization-specific incidence of CRC as observed before the introduction of screening (**Figure 8**).[9]

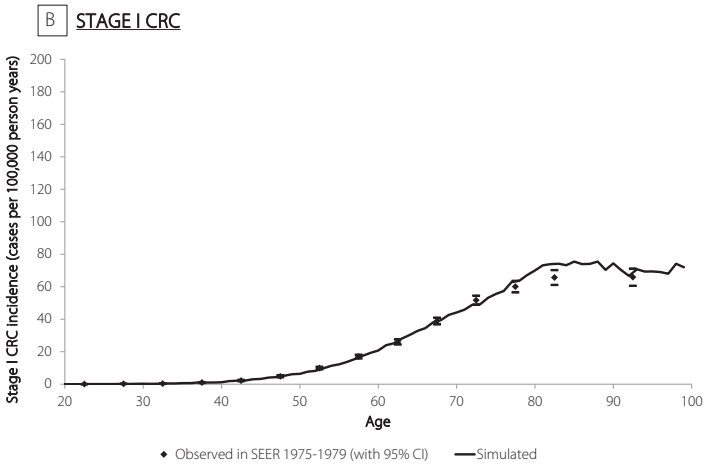
The average durations of the preclinical cancer stages were calibrated to the rates of screen-detected and interval cancers observed in randomized controlled trials evaluating screening using guaiac fecal occult blood tests.[20-22] This exercise has been described extensively in a publication by Lansdorp-Vogelaar and colleagues.[23] The average duration from the emergence of an adenoma until progression into preclinical cancer (i.e. the adenoma dwell-time) was calibrated to the rates of interval cancers (including surveillance detected cancers) observed in a randomized controlled trial evaluating once-only sigmoidoscopy screening (**Figure 9**).[24] We assumed an equal overall dwell-time for adenomas developing into CRC from a medium size (30% of all CRCs) and from a large size (70% of all CRCs). All durations in the adenoma and preclinical cancer phase were drawn from exponential distributions. Durations within the adenoma phase and within the preclinical cancer phase were assumed to be perfectly correlated (i.e. if a small adenoma grows into a medium-sized adenoma rapidly, it will also grow into a large adenoma or develop into CRC rapidly); however, durations in the adenoma phase were assumed to be uncorrelated with durations in the preclinical cancer phase (i.e. a rapidly growing adenoma does not necessarily develop into a rapidly progressing cancer). The proportion of medium sized, non-progressive adenomas growing large and the average duration in the medium size, non-progressive adenoma state were calibrated to size-specific adenoma detection rates observed in a Dutch randomized controlled trial on colonoscopy screening (*data not shown*).

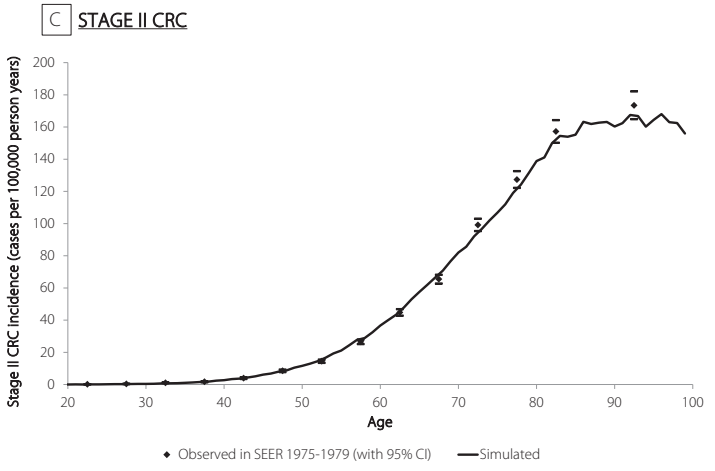
Screening module

Screening will alter some of the simulated life histories: Some cancers will be prevented by the detection and removal of adenomas; other cancers will be detected in an earlier stage with a more favorable survival. As the stage-specific survival of screen-detected CRC as observed in randomized controlled trials on guaiac fecal occult blood testing was substantially more favorable than that of clinically detected CRC, even after correcting for lead-time bias,[23] we assigned those screen-detected cancers that would have been clinically detected in the same stage the survival corresponding to a one stage less progressive cancer. Hence, a cancer screen-detected in stage II, that would also have been clinically diagnosed in stage II, is assigned the survival of a clinically diagnosed stage I cancer. The only exceptions were screen-detected stage IV cancers. These cancers were always assigned the survival of a clinically diagnosed stage IV cancer.

Besides modeling positive health effects of screening, we also model colonoscopy-related complications and overdiagnosis and overtreatment of CRC (i.e. the detection and treatment of cancers that would not have been diagnosed without screening).[25-27]


A TOTAL
Total CRC Incidence (cases per 100,000 person years)
500
450
400
350
300
250
200
150
100
50
0
20
30
40
50
60
70
80
90
100
Age
Observed in SEER 1975-1979 (with 95% CI)
Simulated


B STAGE I CRC
Stage I CRC incidence (cases per 100,000 person years)
200
180
160
140
120
100
80
60
40
20
0
20
30
40
50
60
70
80
90
100
Age
Observed in SEER 1975-1979 (with 95% CI)
Simulated


C STAGE II CRC
Stage II CRC incidence (cases per 100,000 person years)
200
180
160
140
120
100
80
60
40
20
0
20
30
40
50
60
70
80
90
100
Age
Observed in SEER 1975-1979 (with 95% CI)
Simulated

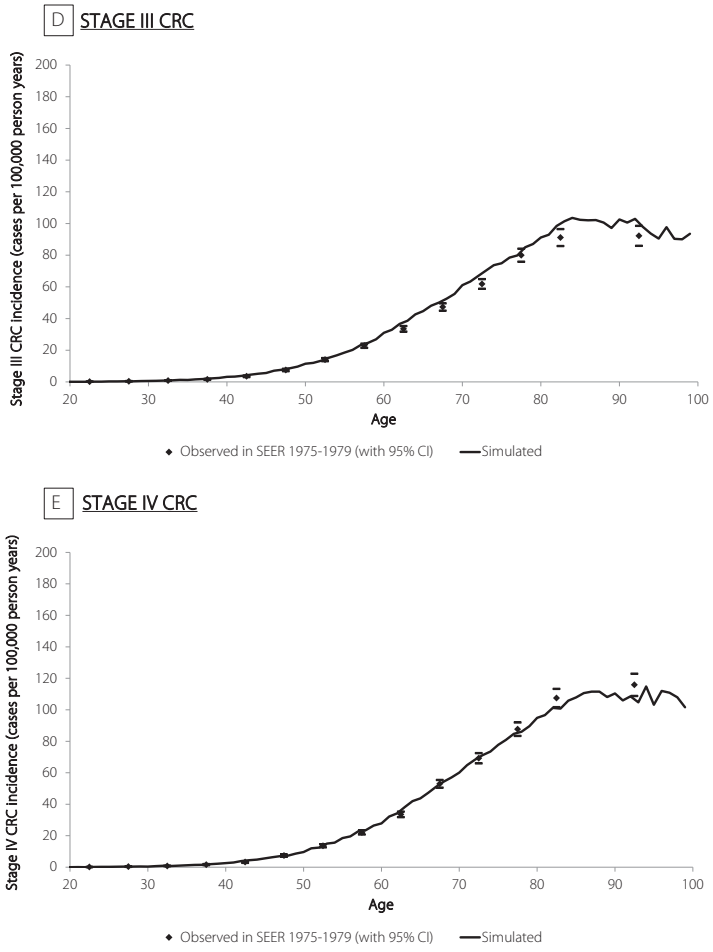


Figure 8. CRC incidence observed before the introduction of screening versus simulated by MISCAN-Colon (total (A), stage I CRC (B), stage II CRC (C), stage III CRC (D), stage IV CRC (E); cases per 100,000 person years).

CRC = colorectal cancer.

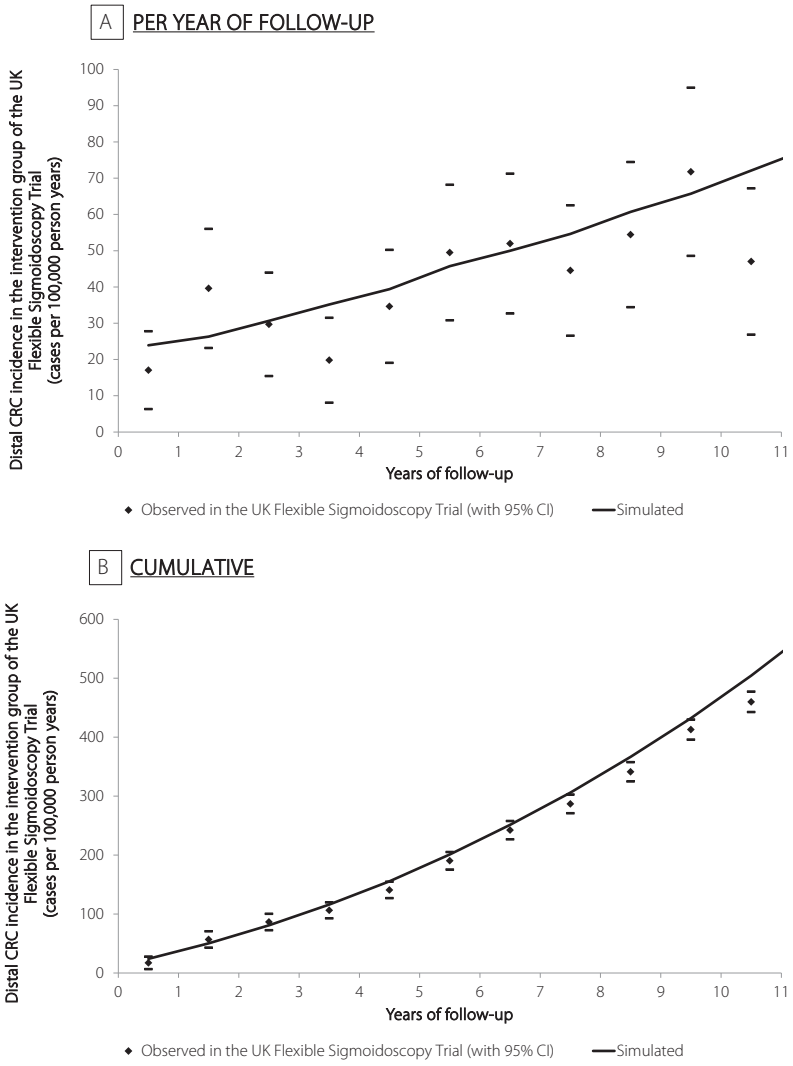


Figure 9. Distal CRC incidence observed in the intervention group of the UK Flexible Sigmoidoscopy trial versus simulated by MISCAN-Colon (per year of follow-up (A), cumulative (B); cases per 100,000 person years).

CRC = colorectal cancer.

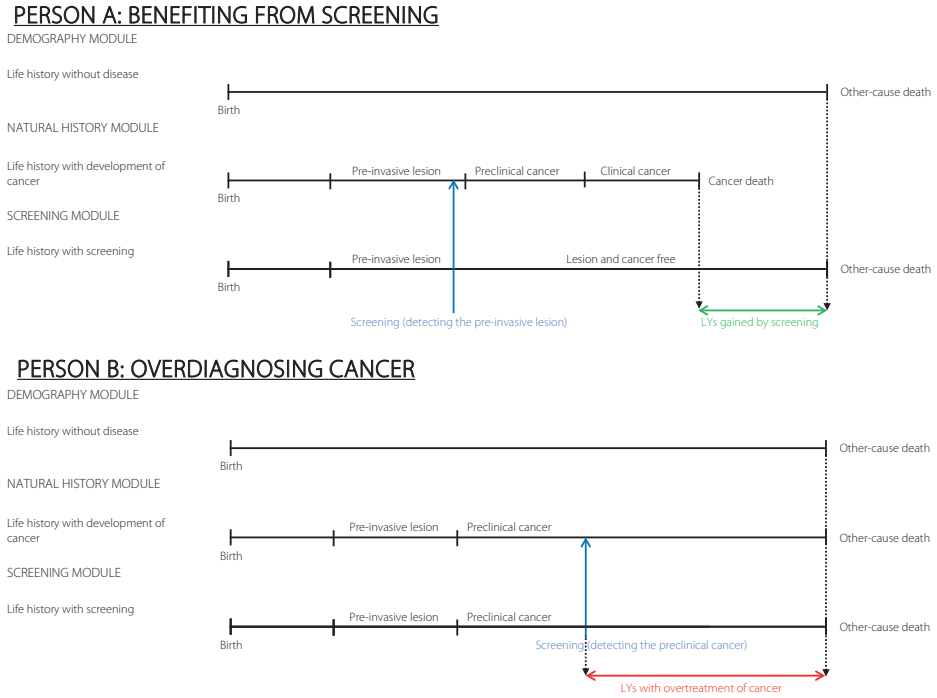


Figure 10. Integrating Modules: Two example Persons.
LYs = life years.

INTEGRATING MODULES

The demography module generates a date of birth and a date of death from other causes for each individual simulated, creating a life history without any lesions or cancer. In Person A in **Figure 10**, the natural history module generates a pre-invasive lesion (e.g. CIN or adenoma). This lesion progresses into preclinical cancer, which is diagnosed because of symptoms and results in cancer death before death from other causes would have occurred. In the screening module a screening examination is simulated, indicated by the blue arrow. During this examination the pre-invasive lesion is detected, and as a result both cancer diagnosis and cancer death are prevented. Hence, in Person A, screening prolongs life by the amount indicated by the green arrow. Person B also develops a pre-invasive lesion, and although this lesion does progress into preclinical cancer, Person B would never have been diagnosed with cancer in a scenario without screening. However, during the screening examination simulated in the screening module, again indicated by the blue arrow, cancer is detected. Hence, in this person screening results in overdiagnosis of cancer: It detects a cancer that would never have been diagnosed in a scenario without screening. Hence, screening does not prolong life, but it does result in additional LYs with cancer care (overtreatment) as indicated by the red arrow.

REFERENCES

1. Statline Database. The Hague: Statistics Netherlands (CBS); 2013.
2. Hospital Diagnosis Statistics 1963-1985. Utrecht: SIG (Information Centre for Health Care), 1985.
3. Netherlands Cancer Registry (NCR). Age-specific and stage-specific cervical cancer incidence (2000-2010) and survival (1989-2009). 2012.
4. Nationwide Network and Registry of Histo- and Cytopathology in the Netherlands (PALGA). PALGA Database for the Evaluation of the Dutch Cervical Cancer Screening Program (PALEBA). 2012.
5. Bulkmans NW, Rozendaal L, Snijders PJ, et al. POBASCAM, a population-based randomized controlled trial for implementation of high-risk HPV testing in cervical screening: design, methods and baseline data of 44,102 women. Int J Cancer 2004; 110(1): 94-101.
6. Lenselink CH, Melchers WJ, Quint WG, et al. Sexual behaviour and HPV infections in 18 to 29 year old women in the pre-vaccine era in the Netherlands. PLoS One 2008; 3(11): e3743.
7. Arias E. United States Life Tables, 2009: Centers for Disease Control and Prevention, 2014.
8. Rutter CM, Johnson EA, Feuer EJ, Knudsen AB, Kuntz KM, Schrag D. Secular trends in colon and rectal cancer relative survival. J Natl Cancer Inst 2013; 105(23): 1806-13.
9. SEER*Stat Database: Incidence–SEER 9 Regs Limited-Use NS-. Surveillance, Epidemiology, and End Results (SEER) Program. 2003. http://www.seer.cancer.gov Accessed July 1st, 2016.
10. Blatt LJ. Polyps of the colon and rectum: Incidence and distribution. Dis Colon Rectum 1961; 4(4): 277-82.
11. Arminski TC, McLean DW. Incidence and Distribution of Adenomatous Polyps of the Colon and Rectum Based on 1,000 Autopsy Examinations. Dis Colon Rectum 1964; 7: 249-61.
12. Bombi JA. Polyps of the colon in Barcelona, Spain. An autopsy study. Cancer 1988; 61(7): 1472-6.
13. Chapman I. Adenomatous polypi of large intestine: incidence and distribution. Ann Surg 1963; 157: 223-6.
14. Clark JC, Collan Y, Eide TJ, et al. Prevalence of polyps in an autopsy series from areas with varying incidence of large-bowel cancer. Int J Cancer 1985; 36(2): 179-86.
15. Jass JR, Young PJ, Robinson EM. Predictors of presence, multiplicity, size and dysplasia of colorectal adenomas. A necropsy study in New Zealand. Gut 1992; 33(11): 1508-14.
16. Johannsen LG, Momsen O, Jacobsen NO. Polyps of the large intestine in Aarhus, Denmark. An autopsy study. Scand J Gastroenterol 1989; 24(7): 799-806.
17. Rickert RR, Auerbach O, Garfinkel L, Hammond EC, Frasca JM. Adenomatous lesions of the large bowel: an autopsy survey. Cancer 1979; 43(5): 1847-57.
18. Vatn MH, Stalsberg H. The prevalence of polyps of the large intestine in Oslo: an autopsy study. Cancer 1982; 49(4): 819-25.
19. Williams AR, Balasooriya BA, Day DW. Polyps and cancer of the large bowel: a necropsy study in Liverpool. Gut 1982; 23(10): 835-42.
20. Hardcastle JD, Chamberlain JO, Robinson MH, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet 1996; 348(9040): 1472-7.
21. Jorgensen OD, Kronborg O, Fenger C. A randomised study of screening for colorectal cancer using faecal occult blood testing: results after 13 years and seven biennial screening rounds. Gut 2002; 50(1): 29-32.
22. Mandel JS, Church TR, Ederer F, Bond JH. Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. J Natl Cancer Inst 1999; 91(5): 434-7.

23. Lansdorp-Vogelaar I, van Ballegooijen M, Boer R, Zauber A, Habbema JD. A novel hypothesis on the sensitivity of the fecal occult blood test: Results of a joint analysis of 3 randomized controlled trials. Cancer 2009; 115(11): 2410-9.
24. Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. Lancet 2010; 375(9726): 1624-33.
25. Warren JL, Klabunde CN, Mariotto AB, et al. Adverse events after outpatient colonoscopy in the Medicare population. Ann Intern Med 2009; 150(12): 849-57, W152.
26. Gatto NM, Frucht H, Sundararajan V, Jacobson JS, Grann VR, Neugut AI. Risk of perforation after colonoscopy and sigmoidoscopy: a population-based study. J Natl Cancer Inst 2003; 95(3): 230-6.
27. van Hees F, Zauber AG, Klabunde CN, Goede SL, Lansdorp-Vogelaar I, van Ballegooijen M. The Appropriateness of More Intensive Colonoscopy Screening Than Recommended in Medicare Beneficiaries - A Modeling Study. JAMA Intern Med 2014; 174(10): 1568-76.

Summary

Screening is the systematic testing of asymptomatic individuals to identify disease or risk factors for disease. This enables the possibility to act earlier, e.g. by starting treatment of the disease earlier. Screening is especially valuable for diseases like cancer, where prognosis is better in earlier stages. In the Netherlands, screening programs exist for early detection of cervical cancer, breast cancer, and colorectal cancer.

Prior to implementation of (either organized or opportunistic) screening, the following three criteria should be met:

1. There should be scientific evidence of screening program effectiveness.
2. The overall benefits of screening should outweigh the harm.
3. There should be scientific evidence of screening program cost-effectiveness.

Both cervical and colorectal cancer are extremely suitable for mass screening. Their natural history is similar in the sense that a long preclinical phase precedes cancer development. During this preclinical phase, benign lesions can be detected by screening. By removing these lesions, cancer development can be averted. Additionally, early detection and treatment of cancer can improve survival significantly.

As with any form of screening, cervical and colorectal cancer screening may be experienced as burdensome, and even come with harm for some of its participants. Whenever possible, attempts should be made to improve the harm-benefit ratio of screening. In this thesis, several opportunities to reduce harms and increase benefits of cervical and colorectal cancer screening were explored. Risk-based screening may be particularly suitable, as in theory, it reduces harms and increases benefits simultaneously.

In general, the optimal screening strategy for a population or risk group is determined in a cost-effectiveness analysis, in which the costs and health benefits of several strategies are compared.

METHODOLOGICAL ISSUES IN COST-EFFECTIVENESS ANALYSES

Whether a strategy is deemed optimal in a cost-effectiveness analysis depends on the set of included strategies for comparison. **Chapter 2** of this thesis has shown that the majority of cost-effectiveness analyses on cervical cancer screening were likely to have omitted relevant strategies from the analysis. In most cases, strategies with a longer screening interval than the one deemed optimal were not included. This likely led to underestimation of the incremental costs per life year gained, and may have led to the identification of suboptimal strategies as being optimal.

REDUCING HARMS

Over the past decades, cytology-based screening has shown to be an effective way of reducing cervical cancer incidence and mortality. Since the discovery that infection with human papillomavirus (HPV) is a necessary cause of cervical cancer, a wide range of HPV tests have been developed. As HPV-based screening provides better protection against invasive cervical cancer, it has already been implemented in the Netherlands (per 1 January 2017), and will soon be implemented in Australia (per 1 December 2017). However, the higher sensitivity of HPV-based screening comes with a reduced specificity, meaning that (false-)positive testing rates will increase. Although for cytology screening, it is known that over-screening (i.e. screening more often than recommended by international guidelines) is neither necessary nor desired, it is still current practice in many countries.

In **Chapter 3** we have estimated the potential harms of switching from cytology to HPV-based screening for women who are over-screened. As with cytology-based screening at an annual or biennial basis it is already unlikely to miss clinically relevant disease, the health benefits of switching to the more sensitive HPV test were small. The potential harms due to increased numbers of (false-)positive tests and referrals for colposcopy, however, were substantial. Altogether, for women who are screened at an annual or biennial basis, switching to HPV-based testing is likely to result in a net loss in health.

Besides psychological distress of cervical cancer screening, treatment of precursor lesions called cervical intraepithelial neoplasia (CIN) may increase risk of subsequent preterm birth. Preterm birth significantly increases the risk of infant morbidity and mortality. To our knowledge, this harm has not been taken into account in (cost-)effectiveness analyses of cervical cancer screening. **Chapter 4** of this thesis provides a first attempt by comparing screening strategies in terms of (quality-adjusted) life years gained by women and (quality-adjusted) life years lost by their future children. Compared to initiating 3-yearly screening at age 30, the benefit of initiating 3-yearly screening at age 21 was limited in terms of life years gained, and even negative in terms of quality-adjusted life years gained. The harms, however, were significant: 223 more preterm births per 100,000 screened women, resulting in 22 more cases of infant morbidity and 6 more infant deaths. In countries like the US, where 3-yearly screening from age 21 is recommended, women with a (future) childwish should therefore consider delaying the start age of cervical cancer screening to age 30.

INCREASING BENEFITS

The benefits of screening can be enhanced by the introduction of new screening modalities. The new screening test could either replace the currently offered test, or could function as an alternative to existing screening methods. The latter could eliminate test-specific barriers that individuals may experience in relation to existing screening methods, and could therefore entice un(der)screened individuals to participate in screening. Whereas replacement of a test requires that the new test is more effective (or at least more cost-effective) than the one currently offered, the addition of an alternative screening method can also be beneficial if the test is less effective.

Over the past 10-15 years, many countries, including the Netherlands, have replaced the use of conventional cytology with that of liquid-based cytology (LBC) tests SurePath and ThinPrep. As compared to conventional cytology, the use of LBC results in fewer samples of unsatisfactory quality. Another advantage of LBC is that it facilitates co-testing (i.e. samples can also be tested on the presence of HPV). Systematic reviews comparing conventional cytology with LBC showed comparable test performance, but assumed no differences between LBC tests. In an earlier study, however, we have shown that in the Netherlands, the use of SurePath was associated with an increased CIN 2+ detection rate compared to conventional cytology, while the use of ThinPrep was not. The analysis presented in **Chapter 5** showed that screening with SurePath also provided better protection against invasive cervical cancer than conventional cytology, while screening with ThinPrep did not.

Since 2017, the program no longer involves primary LBC testing, though. Instead, samples are primarily tested on the presence of oncogenic HPV-types. Women who do not attend screening within 6 months after invitation, are offered HPV self-sampling. In **Chapter 6**, we have shown that this is expected to be (cost-)effective if the relative CIN 2+ sensitivity with a self-collected sample (as opposed to with a clinician-collected sample) is at least 0.95, previously unscreened women are recruited, and total participation increases with at least 6 percentage points. If these requirements are not met, women switching from the regular test to self-sampling may decrease the total effectiveness of the program.

Offering an alternative screening method might also increase adherence with colorectal cancer screening guidelines in the US. The recently updated US Preventive Services Task Force guidelines recommend colorectal cancer screening with one (or two) of six modalities, including also the recently approved multitarget Stool DNA test. Three-yearly screening with this test has now been covered by the Centers for Medicare and Medicaid Services (CMS). In **Chapter 7**, we have estimated whether this screening strategy is a cost-effective alternative to other reimbursed strategies for the Medicare population (i.e. US individuals ages 65 and over), and if not, at what cost, adherence level or frequency of testing it would be. Our results showed that even if adherence with 3-yearly multitarget Stool DNA testing would be 30% higher than with other screening strategies, it would not be cost-effective at its current reimbursement rate of $493. For triennial, biennial or annual multitarget Stool DNA testing, per-test reimbursement rate has to be below $33, $47, or $58, respectively, to be a cost-effective alternative to other reimbursed strategies.

REDUCING HARMS AND INCREASING BENEFITS: OPPORTUNITIES FOR RISK-BASED SCREENING

Risk-based screening provides the opportunity to reduce harms of screening in individuals at low risk for the disease, while increasing benefits in those at higher risk.

Compared to the current screening eligible population, HPV-vaccinated women are expected to be at much lower risk for cervical cancer. Using microsimulation modeling, we found that whereas 8 lifetime screens is optimal for the current (unvaccinated) target population of screening, 3 lifetime screens is optimal for vaccinated women (**Chapter 8**). However, screening tailored to HPV vaccination status might not be feasible due to organizational or ethical matters. If uniform screening is the only viable option, then the intensity of such screening is preferably reduced as vaccination coverage, and thereby herd immunity, increases. We found that when the reduction in HPV prevalence among unvaccinated women is at least half of that in vaccinated women (i.e. herd immunity >50%), then a screening program with 3 lifetime screens becomes more cost-effective for unvaccinated women than one with 8 lifetime screens. Uniform screening should therefore not be adjusted to the risk level in vaccinated women until unvaccinated women benefit from a herd immunity level of at least 50%.

US screening guidelines for colorectal cancer already include more intensive screening for individuals with a positive family history than for people in the general population. Individuals with ≥1 affected first-degree relative, are recommended to start colonoscopy screening earlier (at age 40 instead of age 50) and to repeat screening at shorter intervals (5 years instead of 10 years). In **Chapter 9** we used microsimulation modeling to determine, for every number of affected first-degree relatives, the optimal screening strategy by age. We found that individuals with a positive family history may benefit from intensive screening at young age. However, if individuals consistently have negative colonoscopies, it is unlikely that they are affected by genetic predisposition for colorectal cancer, so continuing intensive colonoscopy screening would provide little or no additional benefit. In such case, the interval can gradually be lengthened with age.

Genetic predisposition to colorectal cancer can also be measured by testing for the presence of common genetic variants that are associated with colorectal cancer risk. With an area under the curve (AUC) of 0.60, the discriminatory performance of current polygenic tests is limited. In **Chapter 10**, we estimated the current and potential future benefit of risk-stratified screening based on polygenic risk, as compared to current uniform screening. We found that at an estimated current price per polygenic test of ~$200, risk-stratified screening is not cost-effective compared to uniform screening. This changes if either the price of the polygenic test drops below $141, or the AUC value of polygenic testing increases beyond 0.65. As population-wide implementation of polygenic testing will likely decrease the price per test, and the expected discovery of more common genetic variants will increase the AUC value of polygenic testing, risk-stratified screening based on polygenic risk may well become cost-effective in the future.

CONCLUSIONS AND RECOMMENDATIONS

Based on the results found in this thesis, we derived the following conclusions:

- Many published cost-effectiveness analyses on cervical cancer screening are likely to have omitted relevant comparator strategies, thereby possibly identifying an inefficient screening strategy as optimal.
- For over-screened women, switching from primary cytology to primary HPV screening slightly reduces cervical cancer risk but comes with a large increase in the number of (false-)positive tests, and can therefore be considered as harmful.
- Cervical cancer screening from age 21 instead of age 30 only slightly reduces cervical cancer risk, while it substantially increases the risk of preterm birth.
- Data from the Dutch cervical cancer screening program suggest that SurePath detects more progressive lesions than conventional cytology, while ThinPrep might detect fewer.
- Offering a self-sample test to non-attendees in a cervical cancer screening program is likely a cost-effective intervention, but its impact does depend on parameters like the relative sensitivity of the self-sample test (as compared to the regular test), the extent to which it generates extra attendance, and the extent to which attendees of regular screening switch to using the self-sample test.
- Although screening with the multitarget stool DNA test is an effective way to reduce colorectal cancer incidence and mortality, the current price of the test makes it a cost-ineffective alternative to other stool-based tests.
- When herd immunity reaches beyond 50%, offering unvaccinated women cervical cancer screening based on the pre-vaccination risk level is no longer cost-effective, and reducing the intensity of uniform screening may be considered.
- It is cost-effective to offer individuals with a family history of colorectal cancer intensive colonoscopy screening from a young age onwards. However, if individuals do not develop any adenomas, it is unlikely that they have a genetic predisposition to colorectal cancer, so continuing intensive screening would provide little or no additional benefit.
- At the discriminatory performance and cost of current polygenic tests, colonoscopy screening based on polygenic risk is not cost-effective compared to uniform screening. This may change with a reduction in the price per polygenic test and with the discovery of more common genetic variants that are associated with colorectal cancer risk.

Based on these conclusions, we formulated the following recommendations:

- In a cancer screening cost-effectiveness analysis, one should include all relevant comparator strategies to obtain reliable cost-effectiveness estimates.
- In screening situations that are characterized by a substantial amount of over-screening, primary cytology screening should not be replaced by primary HPV screening. Reducing the amount of over-screening should be prioritized.

- Cervical cancer screening below age 30 should only be offered after careful evaluation of a woman's expected benefits and harms, taking into account her (future) childwish.
- Since 2017, primary LBC screening is longer performed in the Netherlands, but the fact that the use of SurePath resulted in different screening outcomes than the use of ThinPrep should raise international awareness that the assumed similarity in test characteristics may not be true.
- Although offering an alternative test could increase effectiveness of screening through increasing the uptake among individuals willing to undergo screening, one should be careful with offering alternative tests with inferior characteristics compared to already offered tests because of the potential substitution effect.
- If cervical cancer screening cannot be tailored to HPV-vaccination status, one should wait with adjusting the screening frequency to the risk level in vaccinated women until unvaccinated women benefit from a herd immunity level of at least 50%, i.e. until the reduction in HPV prevalence in unvaccinated women is at least half of that in vaccinated women.
- As screening intensity preferably depends on the disease risk in the target population, herd immunity should be closely monitored in partly vaccinated cohorts.
- Individuals with a constant level of family history of colorectal cancer should consider lengthening their screening interval after several subsequent negative colonoscopies, as this suggests that they are unlikely to have a genetic predisposition for colorectal cancer.
- In addition to discovering more common genetic variants that are associated with colorectal cancer risk, research should aim at identifying comprehensive colorectal cancer risk profiles, including both genetic and environmental risk factors.
- Given high regression rates of CIN 1 and CIN 2, and the potential adverse effects of CIN treatment, future research is needed to estimate the impact of delaying treatment in women with CIN 1 and CIN 2.
- Much of the harm related to cervical and colorectal cancer screening comes from the detection and treatment of non-progressive lesions. Literature suggests the existence of biomarkers that are able to distinguish between progressive and non-progressive lesions. As this may largely reduce the amount of overdiagnosis and overtreatment, more research is needed to identify those biomarkers.

Samenvatting

Screening is het systematisch testen van asymptomatische personen op de aanwezigheid van ziekte of risicofactoren voor ziekte. Dit biedt de mogelijkheid om eerder in te grijpen, door bijvoorbeeld het eerder starten van een behandeling. Screening is bijzonder waardevol voor ziektes als kanker, waarbij de prognose beter is in vroegere stadia. In Nederland bestaan bevolkingsonderzoeken voor de vroege opsporing van baarmoederhalskanker, borstkanker, en darmkanker.

Voorafgaand aan de implementatie van (ofwel georganiseerde, ofwel opportunistische) screening moet worden voldaan aan de volgende drie criteria:

1. Er moet wetenschappelijk bewijs zijn voor de effectiviteit van het screenprogramma.
2. De algehele voordelen moeten groter zijn dan de algehele nadelen.
3. Er moet wetenschappelijk bewijs zijn voor de kosteneffectiviteit van het screenprogramma.

Zowel baarmoederhalskanker als darmkanker zijn uitermate geschikt voor massascreening. Hun natuurlijk beloop is vergelijkbaar in die zin dat een lange preklinische duur voorafgaat aan de ontwikkeling van kanker. Gedurende deze preklinische fase kunnen goedaardige laesies gedetecteerd worden door screening. Door het verwijderen van deze laesies kan de ontwikkeling van kanker worden voorkomen. Daarnaast kan vroege opsporing en behandeling van kanker tot een significante verbetering in overleving leiden.

Zoals bij elke vorm van screening, kan ook screening op baarmoederhalskanker en darmkanker als belastend worden ervaren, en kan het zelfs schadelijk zijn voor sommige deelnemers. Waar mogelijk, zullen pogingen gedaan moeten worden om de voordelen:nadelen ratio te verhogen. In dit proefschrift zijn meerdere mogelijkheden onderzocht om de negatieve effecten van screening op baarmoederhalskanker en darmkanker te verkleinen en de positieve effecten ervan te vergroten. Risicogebaseerde screening kan hierbij uitkomst bieden, aangezien dit in theorie zowel de nadelen verkleint als de voordelen vergroot.

Over het algemeen wordt de optimale screenstrategie voor een bevolking of risicogroep bepaald in een kosteneffectiviteitsanalyse, waarin de kosten en gezondheidseffecten van meerdere strategieën met elkaar worden vergeleken.

METHODOLOGISCHE KWESTIES IN KOSTENEFFECTIVITEITSANALYSES

Of een strategie optimaal wordt bevonden in een kosteneffectiviteitsanalyse hangt af van de set van geïncludeerde strategieën. **Hoofdstuk 2** van dit proefschrift liet zien dat in de meeste kosteneffectiviteitsanalyses van baarmoederhalskankerscreening, relevante strategieën zijn weggelaten uit de analyse. In de meeste gevallen waren er geen strategieën meegenomen met een langer interval dan degene die optimaal werd bevonden. Dit heeft waarschijnlijk geleid tot een onderschatting van de incrementele kosten per gewonnen levensjaar, en zou ertoe geleid kunnen hebben dat suboptimale strategieën als optimaal zijn aangeduid.

VERKLEINEN VAN SCHADELIJKE EFFECTEN

Over de afgelopen decennia heeft cytologie screening laten zien een effectief middel te zijn voor het verlagen van baarmoederhalskankerincidentie en -sterfte. Sinds de ontdekking dat een infectie met het humaan papillomavirus (HPV) een voorwaarde is voor het ontwikkelen van baarmoederhalskanker, is er een veelvoud aan HPV-testen ontwikkeld. Omdat screening op HPV een betere bescherming biedt tegen invasieve baarmoederhalskanker, is dit onlangs in Nederland ingevoerd (per 1 januari 2017) en zal het binnenkort in Australië worden ingevoerd (per 1 december 2017). De hogere sensitiviteit van screening op HPV gaat echter gepaard met een lagere specificiteit, hetgeen betekent dat het aantal (vals-) positieve testuitslagen zal toenemen. Hoewel het voor cytologie screening bekend is dat *over-screening* (d.w.z. frequenter screenen dan wordt aanbevolen in internationale richtlijnen) zowel onnodig als ongewenst is, komt het in veel landen nog regelmatig voor.

In **Hoofdstuk 3** hebben we de mogelijke schade van het overgaan van cytologie op HPV screening geschat voor vrouwen die *over-screened* zijn. Aangezien het bij jaarlijkse of tweejaarlijkse cytologie al onwaarschijnlijk is dat een klinisch relevante laesie gemist wordt, was er slechts een kleine gezondheidswinst bij het overgaan op de sensitievere HPV-test. De mogelijke schade ten aanzien van toegenomen aantallen (vals-)positieve uitslagen en colposcopie verwijzingen was echter aanzienlijk. Alles bij elkaar genomen leidt een overstap van cytologie op HPV screening voor vrouwen die jaarlijks of tweejaarlijks gescreend worden tot een netto verlies in gezondheid.

Naast de psychologische stress van baarmoederhalskankerscreening, kan de behandeling van cervicale intra-epitheliale neoplasie (CIN) tot een verhoogd risico op vroeggeboorte leiden. Vroeggeboorte verhoogt het risico op neonatale morbiditeit en mortaliteit aanzienlijk. Voor zover wij weten zijn er tot op heden geen (kosten)effectiviteitsanalyses van baarmoederhalskankerscreening verricht waarin dit schadelijke effect wordt meegenomen. **Hoofdstuk 4** van dit proefschrift biedt hiertoe een eerste poging, met een vergelijking van screenstrategieën op basis van de (voor kwaliteit van leven gecorrigeerde) gewonnen levensjaren bij vrouwen en de (voor kwaliteit van leven gecorrigeerde) verloren levensjaren bij hun toekomstige kinderen. Vergeleken met het starten van 3-jaarlijkse screening op leeftijd 30, was de opbrengst van het verlagen van de startleeftijd naar 21 jaar beperkt in gewonnen levensjaren, en zelfs negatief in de voor kwaliteit van leven gecorrigeerde levensjaren. De schadelijke effecten waren echter significant: 223 meer vroeggeboortes per 100.000 gescreende vrouwen, resulterend in 22 meer gevallen van neonatale morbiditeit en 6 meer gevallen van neonatale mortaliteit. In landen zoals de VS, waar 3-jaarlijkse screening vanaf leeftijd 21 wordt aanbevolen, zouden vrouwen met een (toekomstige) kinderwens daarom moeten overwegen om baarmoederhalskankerscreening uit te stellen tot leeftijd 30.

VERGROTEN VAN GUNSTIGE EFFECTEN

De gunstige effecten van screening kunnen vergroot worden door de invoering van nieuwe testmethoden. De nieuwe screentest kan ofwel de bestaande test vervangen, ofwel als een alternatief voor huidige screentesten fungeren. Dit laatste zou ervoor kunnen zorgen dat personen die (onlangs) niet hebben deelgenomen aan screening vanwege een aversie tegen bestaande testmethoden, dan wel besluiten zich te laten screenen. Hoewel het bij vervanging van een screentest een vereiste is dat de nieuwe test effectiever (of tenminste kosteneffectiever) is dan de tot dan toe aangeboden test, kan het toevoegen van een test die minder (kosten)effectief is wel voordelen bieden.

Over de laatste 10-15 jaar hebben veel landen, waaronder Nederland, het gebruik van conventionele cytologie vervangen door dat van dunnelaag cytologie (DLC) testen SurePath en ThinPrep. Het gebruik van deze DLC testen resulteert in minder uitstrijkjes van onvoldoende kwaliteit. Een ander voordeel van DLC is dat het zgn. co-testing mogelijk maakt (d.w.z. uitstrijkjes kunnen ook getest worden op de aanwezigheid van HPV). Systematische reviews waarin conventionele cytologie vergeleken is met DLC lieten vergelijkbare testeigenschappen zien, maar veronderstelden geen verschil tussen DLC testen onderling. In een eerdere studie hebben we echter laten zien dat in het Nederlandse bevolkingsonderzoek, in tegenstelling tot het gebruik van ThinPrep, het gebruik van SurePath geassocieerd was met een hoger CIN 2+ detectiecijfer dan conventionele cytologie. De analyse in **Hoofdstuk 5** liet zien dat SurePath ook een betere bescherming bood tegen invasieve baarmoederhalskanker dan conventionele cytologie, terwijl dit niet gold voor ThinPrep.

Sinds 2017 maakt primaire DLC screening echter geen deel meer uit van het bevolkingsonderzoek. In plaats daarvan worden uitstrijkjes primair getest op de aanwezigheid van oncogene HPV-types. Aan vrouwen die niet binnen 6 maanden na uitnodiging deelnemen aan screening wordt een zelfafnameset aangeboden. In **Hoofdstuk 6** hebben we laten zien dat dit naar verwachting (kosten)effectief is als de relatieve CIN 2+ sensitiviteit bij zelf afgenomen materiaal (in vergelijking met door een professional afgenomen materiaal) tenminste 0.95 is, als tot dusver ongescreende vrouwen er gebruik van maken, en als de totale deelname met tenminste 6 procentpunt toeneemt. Als er aan deze voorwaarden niet kan worden voldaan, dan kunnen vrouwen die overstappen van de reguliere test op het gebruik van de zelfafnameset de totale effectiviteit van het programma verlagen.

Het aanbieden van een alternatieve screentest zou ook kunnen leiden tot een betere naleving van de darmkankerscreeningsrichtlijnen in de VS. De recent geüpdatete richtlijnen van de *US Preventive Services Task Force* bevelen darmkankerscreening aan met één (of twee) van zes modaliteiten, waaronder ook de recent goedgekeurde *multitarget Stool DNA test*. Driejaarlijkse screening met deze test wordt inmiddels vergoed door de *Centers for Medicare and Medicaid Services* (CMS). In **Hoofdstuk 7** hebben we berekend of deze screenstrategie een kosteneffectief alternatief is voor andere vergoede screenstrategieën voor de *Medicare* bevolking (d.w.z. 65-plussers in de VS), en zo niet, bij welke kosten, opkomst en screeningsinterval het dat wel zou zijn. Uit onze resultaten bleek dat bij de huidige vergoeding van $493

per test, driejaarlijkse screening met de *multitarget stool DNA test* zelfs niet kosteneffectief zou zijn als de opkomst 30% hoger zou zijn dan bij andere strategieën. Voor driejaarlijkse, tweejaarlijkse en jaarlijkse screening met de *multitarget stool DNA test*, moet de vergoeding per test lager zijn dan respectievelijk $33, $47, of $58, om het een kosteneffectief alternatief te laten zijn voor andere vergoede strategieën.

VERKLEINEN VAN SCHADELIJKE EFFECTEN EN VERGROTEN VAN GUNSTIGE EFFECTEN: MOGELIJKHEDEN VOOR RISICOGEBASEERDE SCREENING

Risicogebaseerde screening biedt de mogelijkheid om de nadelen van screening te verlagen in laag-risico individuen, en tegelijkertijd de voordelen ervan te vergroten in hoger-risico individuen.

Vergeleken met de bevolking die momenteel in aanmerking komt voor screening, hebben HPV-gevaccineerde vrouwen naar verwachting een veel lager risico op baarmoederhalskanker. Met behulp van microsimulatiemodellering vonden we dat waar 8 screenmomenten per leven optimaal is voor de huidige doelgroep van screening, 3 screenmomenten optimaal is voor gevaccineerde vrouwen (**Hoofdstuk 8**). Screening gebaseerd op vaccinatiestatus zou echter niet uitvoerbaar kunnen zijn vanwege organisatorische of ethische kwesties. Als uniforme screening de enige werkbare optie is, dan zal de intensiteit van dergelijke screening teruggebracht moeten worden wanneer de vaccinatiegraad, en daarmee ook de collectieve immuniteit, toeneemt. We vonden dat als de reductie in HPV-prevalentie in ongevaccineerde vrouwen meer dan de helft is van dat in gevaccineerde vrouwen (d.w.z. collectieve immuniteit >50%), dat dan een programma met 3 screenmomenten voor ongevaccineerde vrouwen kosteneffectiever is dan een programma met 8 screenmomenten. Uniforme screening moet daarom niet aangepast worden aan het risiconiveau in gevaccineerde vrouwen totdat ongevaccineerde vrouwen van een collectieve immuniteit genieten van tenminste 50%.

In de VS screeningsrichtlijnen voor darmkanker is voor personen met een positieve familiegeschiedenis al intensievere screening opgenomen dan voor de algemene bevolking. Individuen met ≥1 eerstegraads familielid wordt aanbevolen om eerder te starten met coloscopie screening (op leeftijd 40 in plaats van op leeftijd 50) en om screening met kortere intervallen te herhalen (5 jaar in plaats van 10 jaar). In **Hoofdstuk 9** hebben we microsimulatie modellering gebruikt om, voor elk aantal eerstegraads familieleden, de optimale strategie te bepalen naar leeftijd. We vonden dat individuen met een positieve familiegeschiedenis baat kunnen hebben bij intensieve screening op jonge leeftijd. Echter, als individuen consistent coloscopieën hebben zonder bevindingen, dan is het onwaarschijnlijk dat zij een genetische aanleg hebben voor darmkanker, en biedt het continueren van intensieve coloscopie screening weinig of geen voordelen. In zo'n geval kan het screeningsinterval geleidelijk verlengd worden.

Genetische aanleg voor darmkanker kan ook gemeten worden door te testen op de aanwezigheid van veelvoorkomende genetische varianten die geassocieerd zijn met het risico op darmkanker. Met een *area under the curve* (AUC) van 0.60, is het discriminerend vermogen van huidige polygenetische testen echter beperkt. In **Hoofdstuk 10** hebben we de huidige en mogelijke toekomstige voordelen van risicogestratificeerde screening op basis van polygenetisch risico geschat, ten opzichte van huidige uniforme screening. We vonden dat, bij een geschatte huidige prijs per polygenetische test van ~$200, risicogestratificeerde screening niet kosteneffectiever is dan uniforme screening. Dit verandert wanneer ofwel de prijs van de polygenetische test lager wordt dan $141, ofwel de AUC waarde stijgt tot meer dan 0.65. Aangezien het aanbieden van een polygenetische test op populatieniveau de prijs per test waarschijnlijk zal drukken, en de verwachte ontdekking van meer veelvoorkomende genetische varianten de AUC waarde zal doen stijgen, zou risicogestratificeerde screening in de toekomst wel kosteneffectief kunnen worden.

CONCLUSIES EN AANBEVELINGEN

Op basis van de resultaten in dit proefschrift, kunnen de volgende conclusies worden getrokken:

- Bij veel gepubliceerde kosteneffectiviteitsanalyses van baarmoederhalskankerscreening is het aannemelijk dat relevante strategieën ontbreken, waardoor mogelijk een inefficiënte strategie is aangeduid als zijnde optimaal.
- Voor over-gescreende vrouwen leidt de overgang van primaire cytologie op primaire HPV screening tot een lichte daling in het risico op baarmoederhalskanker, maar dit gaat gepaard met een grote stijging in het aantal (vals-)positieve testen, en kan daardoor als schadelijk beschouwd worden.
- Baarmoederhalskankerscreening vanaf 21 jaar in plaats van vanaf 30 jaar leidt tot een licht verlaagd risico op baarmoederhalskanker, maar verhoogt de kans op vroeggeboorte aanzienlijk.
- Data van het Nederlandse bevolkingsonderzoek baarmoederhalskanker suggereren dat SurePath meer progressieve laesies detecteert dan conventionele cytologie, terwijl ThinPrep er wellicht minder detecteert.
- Het aanbieden van een zelfafnameset aan niet-deelnemers van baarmoederhalskankerscreening is waarschijnlijk kosteneffectief, maar de impact hangt af van parameters als de relatieve sensitiviteit van de zelftest (ten opzichte van de reguliere test), de mate waarin het extra deelname genereert, en de mate waarin vrouwen overstappen van de reguliere test op het gebruik van de zelfafnameset.
- Hoewel screening met de *multitarget Stool DNA test* een effectieve manier is om darmkankerincidentie en –mortaliteit te verlagen, maakt de huidige prijs van de test het een kosten-ineffectief alternatief voor andere ontlastingstesten.

- Wanneer de collectieve immuniteit stijgt tot meer dan 50% is het niet meer kosteneffectief om ongevaccineerde vrouwen baarmoederhalskankerscreening gebaseerd op het pre-vaccinatie risico aan te bieden, en kan een verlaging van de intensiteit van uniforme screening worden overwogen.
- Het is kosteneffectief om individuen met een familiegeschiedenis van darmkanker intensief te screenen vanaf jonge leeftijd. Als zij echter geen adenomen ontwikkelen, dan is het onwaarschijnlijk dat zij een genetische aanleg hebben voor darmkanker en dan biedt het continueren van intensieve screening weinig tot geen voordelen.
- Bij het discriminerend vermogen en de prijs van huidige polygenetische testen is coloscopie screening op basis van polygenetisch risico niet kosteneffectief vergeleken met uniforme screening. Dit kan veranderen door een verlaging van de prijs per polygenetische test of met de ontdekking van meer veelvoorkomende genetische varianten die geassocieerd zijn met het risico op darmkanker.

Op basis van deze conclusies, hebben we de volgende aanbevelingen geformuleerd:

- In een kosteneffectiviteitsanalyse van kankerscreening moeten alle relevante strategieën worden geïncludeerd om een betrouwbare schatting van de kosteneffectiviteit te krijgen.
- In screeningssituaties die gekenmerkt worden door een substantiële mate van over-screening, moet primaire cytologie niet vervangen worden door primaire HPV screening. In zo'n geval moet het terugdringen van de hoeveelheid over-screening geprioriteerd worden.
- Baarmoederhalskankerscreening onder leeftijd 30 zou alleen aangeboden moeten worden na een zorgvuldige afweging van de verwachte voor- en nadelen voor een vrouw, daarbij in acht nemend haar (toekomstige) kinderwens.
- Sinds 2017 wordt primaire DLC screening niet meer uitgevoerd in Nederland, maar het feit dat het gebruik van SurePath tot andere uitkomsten heeft geleid dan het gebruik van ThinPrep zou moeten leiden tot een internationaal bewustzijn dat het onterecht is om geen verschil tussen de testen te veronderstellen.
- Hoewel het aanbieden van een alternatieve test de effectiviteit van screening kan verhogen door het genereren van extra opkomst onder individuen die bereid zijn deel te nemen aan screening, is voorzichtigheid geboden bij het aanbieden van alternatieve testen met inferieure testeigenschappen vergeleken met huidige screentesten vanwege het mogelijke substitutie effect.
- Indien baarmoederhalskankerscreening niet kan afhangen van HPV-vaccinatie status, dan zal men moeten wachten met het bijstellen van het screenprogramma naar het risico in gevaccineerde vrouwen totdat ongevaccineerde vrouwen profiteren van tenminste 50% collectieve immuniteit, d.w.z. totdat de reductie in HPV-prevalentie in ongevaccineerde vrouwen tenminste half zo groot is als die in gevaccineerde vrouwen.

- Aangezien de intensiteit van screening bij voorkeur afhangt van het risico in de doelgroep, moet de collectieve immuniteit nauwkeurig gemonitord worden in gedeeltelijk gevaccineerde cohorten.
- Na meerdere coloscopieën zonder bevindingen, zouden individuen met een onveranderde familiegeschiedenis van darmkanker moeten overwegen om hun screeninginterval te verlengen, aangezien zij waarschijnlijk geen genetische aanleg voor darmkanker hebben.
- Naast het ontdekken van meer veelvoorkomende genetische varianten die geassocieerd zijn met darmkanker zou meer onderzoek gericht moeten worden op het identificeren van volledige risicoprofielen, met daarin zowel omgevings- als genetische risicofactoren.
- Gegeven de hoge mate van CIN 1 en CIN 2 regressie en de mogelijke negatieve effecten van CIN behandeling, is verder onderzoek nodig naar de impact van het uitstellen van de behandeling van vrouwen met CIN 1 en CIN 2.
- Een groot deel van de schadelijke effecten van baarmoederhals- en darmkankerscreening komt voort uit de detectie en behandeling van niet-progressieve laesies. Literatuur suggereert dat er biomarkers bestaan die onderscheid kunnen maken tussen progressieve en niet-progressieve laesies. Aangezien dit de hoeveelheid overdiagnose en overbehandeling in grote mate zou kunnen terugdringen, is meer onderzoek nodig om deze biomarkers te identificeren.

Dankwoord

Het is zover, mijn boekje is af! Zelf dacht ik lange tijd dat er geen beweging in te krijgen was, maar in het laatste jaar belandde alles in een stroomversnelling. Ik kan amper geloven dat het me dan toch gelukt is. Tijd om iedereen, die (direct of indirect) heeft bijgedragen aan de totstandkoming van dit proefschrift, te bedanken!

Allereerst wil ik natuurlijk mijn promotor Harry de Koning en mijn copromotoren Iris Lansdorp-Vogelaar en Inge de Kok heel erg bedanken voor hun bijdrage aan dit proefschrift. Harry, jij was niet direct betrokken bij het schrijven van de afzonderlijke hoofdstukken, maar des te meer bij het samensmelten tot één geheel. Over de samenhang maakte ik mij in het begin zeker zorgen; alle andere promovendi binnen onze sectie deden namelijk onderzoek naar slechts één type kankerscreening. En hoewel Iris van het begin af aan zei dat het wel goed zou komen omdat de kankers "erg op elkaar leken", zag ik dat toch heel anders. Toen we tijdens een voortgangsgesprek na 3,5 jaar tot mijn verbazing niet alleen de gebruikelijke afzonderlijke artikelen bespraken, maar jij over een titel en hoofdstukindeling begon, kwam daar verandering in. Harry, heel erg bedankt voor je waardevolle bijdrage aan dit proefschrift.

Inge, jij hebt mij van het begin af aan onder je hoede genomen. Dat vergde de nodige begeleiding bij het schrijven van mijn eerste artikel. Meerdere malen zorgde jij ervoor dat analyses niet onnodig ingewikkeld waren en dat onze boodschap helder was. Ik wil je heel erg bedanken voor alles wat ik van je heb mogen leren.

Iris, jouw begeleiding begon pas na anderhalf jaar, maar is van onschatbare waarde geweest. Ik wil je heel erg bedanken voor het vertrouwen dat je mij gegeven hebt. Op momenten dat ik me zorgde maakte over of het allemaal nog wel ging lukken, wist jij me gerust te stellen. Ik bewonder je enorm om je intelligentie en betrokkenheid; een begeleider waar menig promovendus jaloers op was.

Dit is ook de plek waar ik Marjolein van Ballegooijen wil bedanken. Hoewel zij niet officieel mijn (co)promotor is, was zij nauw betrokken bij (vrijwel) al mijn publicaties. Marjolein, jouw kritische blik tilde mijn artikelen keer op keer naar een hoger niveau. Dat sommige geniale inzichten pas kwamen als artikelen in mijn ogen al af waren, kon soms wel wat frustrerend zijn, maar eerlijk is eerlijk, het stuk werd er altijd beter van. Het was erg inspirerend om met je te mogen werken, heel erg bedankt!

Een promotie is onmogelijk zonder een promotiecommissie. Veel dank aan Prof.dr. van Kemenade, Prof.dr. C.M. van Duijn en Prof.dr. C.J.L.M. Meijer voor de bereidheid zitting te nemen in mijn promotiecommissie en voor het beoordelen van dit proefschrift. Ik kijk uit naar uw vragen tijdens mijn verdediging.

My colorectal cancer screening publications were products of the CISNET-CRC team. Thank you Ann, for supporting me and encouraging me to present my work at CISNET meetings and international conferences. Thank you Amy, Karen and Carolyn for making the calls and in-person meetings so interesting, I learned so much from you. I would also like to thank all other co-authors for their valuable contributions. And to all who joined our in-person meetings: thanks for a great time!

Op de afdeling maakte ik deel uit van zowel de 'cervix' als de 'colon' groep. Binnen de cervix groep werkten we veel samen aan artikelen en aan opdrachten voor het RIVM. Ook zijn

we meerdere malen samen op congres geweest. Bedankt voor de gezelligheid en de vele interessante discussies Suzette, Kirsten, en Corine, en later ook Heleen en Erik. Binnen de colon groep was er (zeker aan het begin) minder inhoudelijke samenwerking, maar, mede dankzij initiatieven van Iris, des te meer op het sociale vlak; bedankt Luuk, Frank, Alex, Sonja, Miriam, Reinier, Esther en later ook Maaike, Elleke, Amir, Andrea, Dayna en Anne. Dankzij jullie heb ik deel uit mogen maken van één van de meest hechte onderzoeksgroepen van MGZ. Voor mij vormde het skiweekend hierin het absolute hoogtepunt. Lieve Elleke, ik vond het superleuk om dit samen met jou te organiseren, en ben heel erg blij dat je mijn paranimf wilt zijn.

Je dagelijkse collega's bepalen voor een groot deel hoe leuk je werk is. In de afgelopen vierenhalf jaar heb ik een kamer mogen delen met Christiaan, Hilde, Daan, Miriam, Ingrid, Roelie, Elise Peters, Tjeerd, Marleen, Timor, Marieke, Nikki, Elise van Beeck, Astrid, Isabelle, Pericles en Famke. Ik ben maar zelden met tegenzin naar mijn werk gegaan. Heel erg bedankt voor jullie betrokkenheid, en de vele gezellige lunches en etentjes. De Escape Rooms met Na-2123 en de Efteling met Na-2309 waren beide onvergetelijk.

Naast werk is het ook nodig om zo nu en dan te ontspannen en leuke dingen te doen. Gelukkig waren er vaak genoeg drankjes, etentjes, en uiteraard de jaarlijkse weekendjes weg met de vriendengroep. Ook de avondjes in Vrienden waren soms onvergetelijk. Van iedereen is er één iemand die ik bij naam wil noemen; lieve Jill, ik vind het superleuk dat je mijn paranimf wilt zijn!

Tenslotte wil ik natuurlijk mijn familie bedanken. Lieve mam, heel erg bedankt voor je onvoorwaardelijke liefde en vertrouwen in mij. Het is zo leuk om te zien hoe trots jij op mij bent, en dat terwijl je zelf helemaal niets van die nare bevolkingsonderzoeken moet hebben. Jan Roelf, mijn tweede vader, bedankt voor je onaflatende belangstelling. Het was heel fijn om samen te praten over niet alleen mijn onderzoek, maar over van alles en nog wat. Lieve Goof en Nan, grote broer en zus, jullie zijn altijd een voorbeeld voor mij geweest. Ik weet niet wat ik zonder jullie zou moeten. Lieve pap, ik vind het geweldig dat je speciaal voor mijn promotie naar Nederland komt. Wij komen snel weer een keer jouw kant op!

Natuurlijk wil ik hierbij ook graag van de gelegenheid gebruik maken om mijn schoonouders te bedanken; ik heb mij van het begin af aan erg welkom gevoeld. Jullie zorgzaamheid en vrijgevigheid beschouw ik zeker niet als vanzelfsprekend.

En dan als allerlaatste, lieve Tim, heel erg bedankt voor alle steun die je me gegeven hebt. Wanneer ik de moed verloor, was jij degene die mij aanspoorde om door te gaan. Wanneer ik ergens niet uit kwam, was je altijd bereid om met me mee te denken. Maar misschien nog wel het belangrijkste, was dat jij ervoor zorgde dat ik goed voor mezelf bleef zorgen, en dat ik ook af en toe tijd nam om te ontspannen, iets dat ik met name naar het einde toe hard nodig had! Bedankt voor alle mooie jaren die we samen gehad hebben. Ik hoop dat er nog vele zullen volgen.

About the author

CURRICULUM VITAE

Steffie Katinka Naber was born on the 25th of March 1988, in Zoetermeer, the Netherlands. In 2006, she completed her secondary education at the Stedelijk College in Zoetermeer. That same year she started studying 'Economics and Business Economics' at the Erasmus University in Rotterdam. After completing the first year, she switched to the more mathematical program 'Econometrics and Operations Research'. In 2010 she did an internship at the head office of the Dutch Railways in Utrecht, where she studied the forecasting of future conflicts in the personnel and material schedule due to current delays of trains. She finished her bachelor degree with a thesis on the appropriateness of different algorithms for routing problems with stochastic demand. Afterwards she continued with the master degree 'Econometrics and Management Science', majoring in 'Operations Research and Quantitative Logistics'. This included an internship at the department of Sustainable Transport and Logistics at the Netherlands Organisation for applied scientific research TNO in Delft. Her master's thesis was about different game theoretical methods to allocate CO_2 emission to customers on a distribution route, of which a scientific article was published in Omega in 2015.

Since August 2012 she is working at the Department of Public Health of the Erasmus University Medical Center in Rotterdam. At first, her research focused on the (cost-)effectiveness of cervical cancer screening in the Netherlands. In 2014 she also began working on projects that aim to improve the (cost-)effectiveness of colorectal cancer screening in the United States, Canada, and Australia. This included using the MISCAN microsimulation model to inform the National Institute for Public Health and the Environment (RIVM) and the United States Preventive Services Task Force (USPSTF) on the expected harms and benefits of several screening strategies.

LIST OF PUBLICATIONS

This thesis

O'Mahony JF, **Naber SK**, Normand C, Sharp L, O'Leary JJ, de Kok IMCM. Beware of Kinked Frontiers: A Systematic Review of the Choice of Comparator Strategies in Cost-Effectiveness Analyses of Human Papillomavirus Testing in Cervical Screening. Value Health 2015;18:1138-51.

Naber SK, de Kok IMCM, Matthijsse SM, van Ballegooijen M. The potential harms of primary human papillomavirus screening in over-screened women: a microsimulation study. Cancer Causes Control 2016;27:569-81.

Rozemeijer K, **Naber SK**, Penning C, Overbeek LIH, Looman CWN, de Kok IMCM, Matthijsse SM, Rebolj M, van Kemenade FJ, van Ballegooijen M. Cervical cancer incidence after normal cytological sample in routine screening using SurePath, ThinPrep, and conventional cytology: population based study. BMJ 2017;356:j504.

Rozemeijer K, de Kok IMCM, **Naber SK**, van Kemenade FJ, Penning C, van Rosmalen J, van Ballegooijen M. Offering Self-Sampling to Non-Attendees of Organized Primary HPV Screening: When Do Harms Outweigh the Benefits? Cancer Epidemiol Biomarkers Prev 2015;24:773-82.

Naber SK, Matthijsse SM, Rozemeijer K, Penning C, de Kok IMCM, van Ballegooijen M. Cervical Cancer Screening in Partly HPV Vaccinated Cohorts - A Cost-Effectiveness Analysis. PLoS One 2016;11:e0145548.

Other publications

Matthijsse SM, Hontelez JA, **Naber SK**, van Rosmalen J, Rozemeijer K, Penning C, Bakker R, van Ballegooijen M, de Kok IMCM, de Vlas SJ. The estimated impact of natural immunity on the effectiveness of human papillomavirus vaccination. Vaccine 2015;33:5357-64.

Rozemeijer K, van Kemenade FJ, Penning C, Matthijsse SM, **Naber SK**, van Rosmalen J, van Ballegooijen M, de Kok IMCM. Exploring the trend of increased cervical intraepithelial neoplasia detection rates in the Netherlands. J Med Screen 2015;22(3):144-50.

Rozemeijer K, de Kok IMCM, **Naber SK**, van Kemenade FJ, Penning C, van Rosmalen J, van Ballegooijen M. When Is It Effective to Offer Self-Sampling to Non-Attendees-Response. Cancer Epidemiol Biomarkers Prev 2015;24:1296.

Rozemeijer K, Penning C, Siebers AG, **Naber SK**, Matthijsse SM, van Ballegooijen M, van Kemenade FJ, de Kok IMCM. Comparing SurePath, ThinPrep, and conventional cytology as primary test method: SurePath is associated with increased CIN II(+) detection rates. Cancer Causes Control 2016;27:15-25.

van Luijt PA, Rozemeijer K, **Naber SK**, Heijnsdijk EA, van Rosmalen J, van Ballegooijen M, de Koning HJ. The role of pre-invasive disease in overdiagnosis: A microsimulation study comparing mass screening for breast cancer and cervical cancer. J Med Screen 2016;23(4):210-216.

Matthijsse SM, Hontelez JAC, **Naber SK**, Rozemeijer K, de Kok IMCM, Bakker R, van Ballegooijen M, van Rosmalen J, de Vlas SJ. Public health benefits of routine human papillomavirus vaccination for adults in the Netherlands: a mathematical modeling study. J Infect Dis 2016;214(6):854-61.

Knudsen AB, Zauber AG, Rutter CM, **Naber SK**, Doria-Rose VP, Pabiniak C, Johanson C, Fischer SE, Lansdorp-Vogelaar I, Kuntz KM. Estimation of Benefits, Burden, and Harms of Colorectal Cancer Screening Strategies. Modeling Study for the US Preventive Services Task Force. JAMA 2016;315(23):2595-609.

Naber SK, de Ree DA, Spliet R, van den Heuvel W. Allocating CO_2 emission to customers on a distribution route. Omega 2015;54:191-9.

PHD PORTFOLIO

Name PhD student:	Steffie K. Naber
Erasmus MC department:	Public Health
PhD period:	2012-2016
Promotor:	Prof.dr. H.J. de Koning
Supervisors:	Dr. I. Lansdorp-Vogelaar, Dr. I.M.C.M. de Kok

	Year	Workload (ECTS)
PhD training		
General academic courses		
Academic Writing	2013	1.0
Integrity in Research	2014	0.3
Individual Supervision	2016	0.1
Courses at the Netherlands Institute for Health Sciences (NIHES)		
Quality of Life Measurement (HS11)	2013	0.9
Primary and Secondary Prevention Research (ESP45)	2013	0.7
Planning and Evaluation of Screening (HS05)	2014	1.4
Oral Presentations		
Presentations at the Department of Public Health, Erasmus MC, Rotterdam, the Netherlands	2012 – 2016	1.5
Presentations at meetings of the Cancer Intervention and Surveillance Modeling Network (CISNET), United States	2014 – 2016	3.0
Teleconference presentation at a meeting of the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group	2014	0.6
International Cancer Screening Network (ICSN), Rotterdam, the Netherlands	2015	0.6
Eurogin, Sevilla, Spain	2015	0.6
Eurogin, Salzburg, Austria	2016	0.6
Poster Presentations		
Society for Medical Decision Making (SMDM), Baltimore, MD, United States	2013	0.3
Human papillomavirus (HPV), Seattle, WA, United States	2014	0.3
American Society for Human Genetics (ASHG), San Diego, CA, United States (2x)	2014	0.6
Human papillomavirus (HPV), Lisbon, Portugal	2015	0.3
International Cancer Screening Network (ICSN), Rotterdam, the Netherlands (2x)	2015	0.6
Digestive Disease Week (DDW), San Diego, CA, United States	2016	0.3
Conferences		
Society for Medical Decision Making (SMDM), Baltimore, MD, United States	2013	1.4

Human papillomavirus (HPV), Seattle, WA, United States	2014	1.7
American Society for Human Genetics (ASHG), San Diego, CA, United States	2014	1.1
Eurogin, Sevilla, Spain	2015	1.1
World Endoscopy Organization (WEO), Washington, DC, United States	2015	0.3
International Cancer Screening Network (ICSN), Rotterdam, the Netherlands	2015	0.9
Human papillomavirus (HPV), Lisbon, Portugal	2015	1.3
World Endoscopy Organization (WEO), San Diego, CA, United States	2016	0.3
Digestive Disease Week (DDW), San Diego, CA, United States	2016	1.1
Eurogin, Salzburg, Austria	2016	1.0
Seminars and symposia		
Seminars at the Department of Public Health, Erasmus MC, Rotterdam, the Netherlands	2012 – 2016	3.1
Qiagen Symposium, Utrecht, the Netherlands	2013	0.1
Nederlandse Associatie voor Community Genetics en Public Health Genomics (NACGG) meeting, Utrecht, the Netherlands	2015	0.1
Nederlands Vereniging voor Oncologie (NVvO) meeting, Amsterdam, the Netherlands	2016	0.3

Review activities

American Journal of Managed Care (AJMC)	2015	0.3
PLoS One	2016	0.2
Value in Health	2016	0.2

Teaching

Correcting Bachelor essays of 3rd year Medical Students	2013 – 2016	4.3
Supervising Community projects	2015, 2016	1.4
Lecturing "Genomics and Screening" as part of the course Planning and Evaluation of Screening (NIHES, HS05)	2015, 2016	1.8